Construction of a duck whole genome radiation hybrid panel : an aid for NGS whole genome assembly and a contribution to avian comparative maps

Man Rao

Université de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :
Université Toulouse III Paul Sabatier (UT3 Paul Sabatier)

Discipline ou spécialité :

Genetics

Présentée et soutenue par :
Man RAO

le : vendredi 19 octobre 2012

Titre :

Construction of a duck whole genome radiation hybrid panel: an aid for NGS whole genome assembly and a contribution to avian comparative maps

---

Ecole doctorale :
Biologie, Santé, Biotechnologies (BSB)

Unité de recherche :
Laboratoire Génétique Cellulaire, INRA

Directeur(s) de Thèse :

Dr. Alain VIGNAL

Rapporteurs :

Dr. Darren GRIFFIN
Dr. Christophe HITTE

Membre(s) du jury :

Dr. David CRIBBS,	Professeur de l'Université Paul Sabatier	Président
Dr. Christophe HITTE,	Ingénieur de Recherche à CNRS	Rapporteur
Dr. Darren GRIFFIN,	Professeur de l'Université Kent (UK)	Rapporteur
Dr. Bertrand BED'HOM,	Ingénieur de Recherche à l'INRA	Examinateur
Dr. Alain VIGNAL,	Directeur de Recherche à l'INRA	Examinateur

Acknowledgement

This dissertation would not have been possible without the guidance and the help of all the kind people surrounding me.

First of all, many thanks to the Chinese Scholarship Council and the INRA Département Génétique Animale for providing financial support during my stay at the UMR Génétique Cellulaire at INRA Toulouse.

I would like to thank all members of the jury for evaluating my work. I am very grateful that Dr. Chistophe Hitte and Dr. Darren Griffin have accepted to be the referees and take the time to read my dissertation.

Thanks to members of my comité de these, Dr. Thomas Faraut, Dr. Bertrand Bed'hom, Dr. Christian Diot and Dr. Francis Galibert, for guiding me and giving me useful advice during the course of my project.

I appreciate that Dr. Philippe Mulsant, our ex-director, having his farewell party just before I wrote this section, and Dr. Martine Yerle have, accepted me as a member of this laboratory. Enjoy your brand new life, Philippe! Martine, thanks for helping me a lot both in my personal life and work.

Foremost, I would like to express my sincere gratitude to my supervisor, Dr. Alain Vignal, the best mentor I have ever met! There are too many things I want to thank for, but in a word, he is my icon!

Dr. Mireille Morisson, a very lovely and well-organized woman full of passion, I am thankful for her sincerity and encouragement. She made a lot of efforts to help me adapting the life in

France, assisted my work even though she was not always in good condition, and helped me getting through many administrative problems and so on.

Many thanks to Dr. Thomas Faraut, a “philosopher”-like researcher who always welcomes newcomers. I enjoy and have learned a lot by working with him. I was a dummy for bioinformatics and biomath, although I still am. He is very patient to explain the equations, probabilities, distributions and programming for me.

Thanks a lot to Dr. Valérie Fillon, a wonderful and very kind woman, sharing the same office with me and very familiar with administrative procedures. I like discussing all the topics with her. Cited a saying from her, I would like to be her cow.

I am thankful for the help from Katia Fève, Emmanuel Labarthe, Sophie Leroux and Emeline Lhuillier. They provide so many assistances for my project and introduce me some new technologies, and they always try to communicate with me in English.

Mille merci à Chantal Delcros et Suzanne Bardes. Vous êtes très gentilles, avez offert beaucoup d’assistance technique et de cours de Français dans le labo. J’ai passé du temps avec vous dans la salle de culture, le coin café et dans les couloirs. J’aime bien parler avec vous sur la culture Française, la cuisine, les vacances, la gastronomie et sur tout autre sujets. Vous me manquerez beaucoup après mon depart.

I am grateful that Dr. Ning Li and Dr. Xiaoxiang Hu have offered to me this opportunity of studying in France.

Thanks for help, advices and encouragements from Dr. Magali SanCristobal, Dr. Christele Marie-Etancelin, Dr. Frédéric Pitel, Dr. Hervé Acloque, Dr. Julie Demars, Manuela Ferre and Florence Comayras.

Best wishes to Maria-Ines Fariello, Laure Fresard, Sarah Rousseau, Yoannah François and Annabelle Congras, and “bon courage” for all of you.

I really enjoyed these 4 years in the lab. It is like a big family to me and I always feel free and comfortable to get help from. Thanks a lot a lot to all the members in the lab.

I would like to thank to other people that I have met in France: Ojo, Sophie Ruzafa-Serreno, Natacha Nikolic, Diane Equerre, Margarita Cano, Yuna Blum, Laure Huge, Siham Ouchia.

Also, I would like to thank my sweet friends Lumei Wei, Suhua Guan, Yanling Chen, Shu li, Zheya Sheng, Lei Yang and Binbin Zhou. Thank you so much that we can meet here and get along like a family. Especially to my cute roommates, I am so glad that we can live, share our life and take care of each other together. Meanwhile, I also would like to say “thank you” to my friends in China for everything even though we are far from each other: Zhengyuan Zhai, Fang Yuan, Chendong Guo, Yanqiang Wang, Chungang Feng,Xiaorong Gu and Qianghe Li.

At the end, I would like to thank my family: my parents, my younger sister and my husband. They are always very supportive. Especially thanks to my sis and my husband for taking over my responsibilities to look after my parents while I was away from my country.

I deeply appreciate all the people mentioned above and some people I may have forgotten to mention, for being kind and supportive to me. Thanks again and all the best for all of you.

Construction of Duck whole genome radiation hybrid panel: an aid for NGS whole genome assembly and a contribution to avian comparative maps

UMR444 Génétique Cellulaire, 24 chemin de borde-rouge, BP52627, Castanet-Tolosan, 31326
Directeur de thèse : Alain VIGNAL

Duck is a very important agronomic species in France, especially for fatty liver industry which presents 75% worldwide production. Moreover, duck is also a scientific model for avian influenza research as it is a natural reservoir for avian influenza viruses. The work presented here is part of the international collaboration on duck genome sequencing, including SNP detection and mapping, EST sequencing. Our goal is to provide a genome map allowing for fine mapping QTL and identifying candidate genes involved in expression of agronomic traits.

A panel composed of 90 radiation hybrids was produced by fusing irradiated duck donor cells with hamster cells. To avoid large-scale culture of the clones, PCR genotyping involving Whole Genome Amplification (WGA) and/or reduction of reaction volumes were tested and two first maps for duck chromosomes were made. We also used the PCR genotyping method to test for the quality of duck sequence scaffold assemblies, which had been produced by the Beijing Genome Institute (BGI, China). Finally, to cover the whole genome, we performed a low-pass sequencing (0.1X depth) of hybrids, allowing for rapid map development. These maps allow the detection of chromosomal rearrangements that have taken place between the duck and chicken genomes, which have diverged 80 million years ago.

Keywords: RH mapping, duck genome assembly, comparative genomics, parallel sequencing.

Le canard est une espèce d'importance agronomique en France, principalement à travers l'industrie de foie gras, qui représente plus de 75% de la production mondiale. De plus, c'est aussi un modèle important pour l'étude de l'infection par le virus influenza, pour lequel les oiseaux aquatiques sont un réservoir naturel, car porteurs asymptomatiques. Les travaux réalisés lors de la thèse se situent dans le contexte international de l'étude du génome du canard, comportant la séquence du génome, le séquençage d'EST et l'identification et la cartographie de SNP. Le but à terme pour l'INRA étant de disposer des connaissances sur le génome nécessaires pour la cartographie fine de QTL et l'identification de gènes impliqués dans l'expression de caractères agronomiques.

Un panel de 90 d'hybrides irradiés (panel RH) a été réalisé par fusion de cellules donneuses de canard irradiées avec des cellules receveuses de hamster. Afin d'éviter la culture à grande échelle des clones cellulaires, des méthodes de génotypage par PCR utilisant l'amplification complète du génome (WGA) et/ou la réduction des volumes réactionnels ont été testées et deux premières cartes de chromosomes ont ainsi été réalisées. Nous avons également utilisé le génotypage par PCR pour vérifier la qualité de l'assemblage des scaffolds du génome du canard, réalisés par séquençage nouvelle génération Illumina au Beijing Genome Institute (BGI, Chine). Finalement, afin de couvrir le génome complet, nous avons entrepris un séquençage léger (0,1X de profondeur) d'hybrides, permettant une réalisation de cartes plus rapides que par PCR. Ces cartes permettent la détection des réarrangements chromosomiques existant entre les génomes de la poule et du canard, qui sont distants de 80 millions d'années.

Mots clés : carte d'hybrides irradiés, assemblage du génome du canard, génomique comparée, séquençage parallèle.

Table of content

List of figures

List of abbreviations

APS: **A**denosine 5'-**p**hospho**s**ulfate

ATP: **A**denosine **t**riphos**p**hate

bp: **b**ase **p**air

BAC: **B**acterial **A**rtificial **C**hromosome

BGI: **B**eijing **G**enomics **I**nstitute

BTA: **B**enzene-1,3,5-**t**riacetic **A**cid

CAU: **C**hina **A**gricutural **U**niversity

cM: **c**enti**m**organ

CR1: **C**hicken **R**epeat 1

ddNTP: dideoxyribonucleic acid

dNTP: dideoxyribonucleic acid

DOP: **D**egenerate **O**ligonucleotide **P**CR

EST: **E**xpression **S**equence **T**ag

FAO: **F**ood and **A**griculture **O**rganization

FISH: **F**luorescent ***in situ*** **H**ybridization

GGA : Chicken chromosome (***G****allus* ***Ga****llus*)

GWAS: **G**enome **W**ide **A**ssociation **S**tudy

HA : **H**em**a**gglutinin

HAT: **H**ypoxanthine **A**minopterin **T**hymidine

HPAI: **H**ighly **P**athogenic **A**vian **I**nfluenza

HPRT: **H**ypoxanthine-guanine **P**hospho**r**ibosyl**t**ransferase

HSA: Human chromosome (***H****omo* ***Sa****pien*)

IFGT: **I**rradiation and **F**usion **G**ene **T**ransfer

INRA: **I**nstitut **N**ational de la **R**echerche **A**gronomique

IRS: **I**nterspersed **R**epetitive **S**equence

LG: **L**inkage **G**roup

LINE: **L**ong **I**nterspersed **N**ucleotide **E**lement

LPAI: **L**ow **P**athogenic **A**vian **I**nfluenza

LSD : **L**ineage-**S**pecific **D**uplication

LTR: **L**ong **T**erminal **R**epeat

MPSS: Massively Parallel Signature Sequencing

MDA: **M**ultiple **D**isplacement **A**mplification

MGA : Turkey chromosome (***M****eleagris* ***Ga****llopavo*)

MHC : **M**ajor **H**istocompatibility **C**omplex

MMU : Mouse chromosome (***M****us* ***Mu****sculus*)

MYA : **M**illion **Y**e**a**rs

NA : **N**eur**a**minidase

NGS: **N**ext **G**eneration **S**equencing

PAPI : **P**oly(**A**) **p**olymerase **I**

PEP: **P**rimer **E**xtension **P**re-amplification

PFGE: **P**ulse-**F**ield **G**el **E**lectrophoresis

PTP: **P**ico**T**iter**P**late

qPCR: quantitative PCR

QTL: **Q**uantitave **T**rait **L**oci

RFLP: **R**estriction **F**ragment **L**ength **P**olymorphism

RH: **R**adiation **H**ybrid

RIG-I: **R**etinoic acid **I**nducible **G**ene **I**

IFN: **I**nter**f**ero**n**

RNO: Rat chromosome (***R****attus* ***No****vegicus*)

SD: **S**egmental **D**uplication

SINE: **S**hort **I**nterspersed **N**ucleotide **E**lement

SNP: **S**ingle **N**ucleotide **P**olymorphism

SSLP: **S**imple **S**equence **L**ength **P**olymorphism

TGS : **T**hird **G**eneration **S**equencing

TK: **T**hymidine **K**inase

tSMS: **t**rue **S**ingle **M**olecule **S**equencing

VNTR: **V**ariable **N**umber of **T**andem **R**epeats

VT: **V**irtual **T**erminator

WGA: **W**hole **G**enome **A**mplification

WGRH: **W**hole **G**enome **R**adiation **H**ybrid

ZMW: **Z**ero-**M**ode **W**aveguide

Chapter I

General Introduction

1. General information on ducks

1.1 Taxonomy & Domestication

Duck is the common name for a number of species in the Anatidae family of Anseriforms. The ducks are divided between several subfamilies in the Anatidae family; they do not represent a monophyletic group but form a taxon, since swans and geese are not considered ducks.

All domestic ducks descent from the characteristically green-headed wild mallard, *Anas platyrhynchos,* except for the Muscovy duck (*cairina moschata*). The name comes from the Latin anas (a duck) and a combination of two Greek words, platus (broad) and rhynchos (bill). The Muscovy duck is larger than the Mallard in size and was domesticated by South American Indians long before Europeans arrived on the continent.

No one knows for certain when Mallards were first domesticated, but there is some evidence to suggest that Egyptians used ducks in religious ceremonies around 1,353 B.C and possibly also bred them for food. Paintings and carvings in the tomb at Saqqara and "Astronomer to Amun" at the Karnak Temple in Egypt show that more than 3,000 years ago migratory wildfowl were hunted and trapped with large, hexagonal-shaped clap-nets in the extensive swamplands of the Nile delta. These ducks were kept in large aviaries and were force-fed before slaughtering to provide a ready supply of meat throughout the year.

The Southeast Asians were also raising ducks in captivity prior to 500 B.C. But there are some reports suggesting that domestication of duck occurred about 4,000 years ago in China. Wucheng suggests that pottery ducks excavated in the Yan Shi Menkou Mountain in Fujian Province (south China) provides evidences that domestication of duck may have occured during the New Stone Age between 4,000 and 10,000 years ago (Wucheng 1988).

Clayton described a report in the Chinese literature by Yeh, who investigated archaeological evidence and suggested ducks were domesticated in China at least 3,000 years ago (Clayton 1984).

Clayton suggests that the history of domestication of the common duck in both China and Western Europe is obscure but the range of types emanating from the Far East suggests South-east Asia as a major centre of domestication (Clayton 1984). The archaeological evidence along with a favourable environment and agriculture suggest that ducks were probably domesticated in southern China at least 1,500 years before they were separately domesticated in Western Europe (Cherry and Morris 2008).

1.2 Natural habitat and habits

Ducks have a cosmopolitan distribution occurring across most of the world except for Antarctica. A number of species manage to live on sub-Antarctic islands like South Georgina and the Auckland Islands. Many ducks have managed to establish themselves on oceanic islands such as Hawaii, New Zealand and Kerguelen, although many of these species and populations are threatened or have become extinct.

Ducks are mostly aquatic birds, usually smaller than the swans and geese, and may be found in both fresh water and sea water. Ducks exploit a variety of food sources such as grasses, seeds, aquatic plants, fish, insects, small amphibians, worms, and small mollusks. Their natural diet is normally about 90% vegetable matters (seeds, berries, fruits, nuts, bulbs, roots, succulent leaves, and grasses) and 10% animal matter (insects, snails, slugs, leeches, worms, eels, crustacean, and an occasional small fish or tadpole). They have little ability to utilize dietary fiber. Although they eat considerable quantities of tender grass, they are not true grazers like geese, and don’t eat coarse grass and weeds at all. Sand and gravel are swallowed to serve as “grindstones” in the gizzard.

Ducks usually have a long lifespan in natural condition and it is not rare that a duck can live for up to 8 years; there are some reports of exceptional ducks living more than 20 years.

Despite domestication and selection over perhaps more than 3,000 years, domesticated duck still shares many similarities with wild Mallard. Incubation takes about 28 days except for Muscovy duck which needs 35 days, commencing after the last egg is laid to enable all ducklings to hatch more or less together. Young ducks fledge by about 50 days and achieve adult maturation live weight at about 12-14 weeks of age by which time feathering with maturation of primary and secondary wing feather is complete (Cherry and Morris 2008).

Despite large differences in size, color and appearance, all the domesticated duck breeds derived from the Mallard can interbreed freely and produce fertile offspring. Depending on the breed (and season), a female may reach sexual maturity at about 20 weeks of age. Most begin laying at 20-26 weeks, but the best laying varieties start at 16-18 weeks and lay profitably for 2 years.

Ducks are very efficient at converting diet into meat and egg, meaning that duck has a very high feeding efficiency. The most common domesticated duck breed hitherto is named Beijing duck, which is the most popular and major meat-type breed. They can convert 2.4-2.6 kg of concentrated feed into 1 kg of weight gain in confined conditions. The only domestic animal that has higher feeding efficiency is the broiler chicken (Cherry and Morris 2008).

Ducks are adapted to environments with humid climates, such as wetland, swamplands, rivers, lakes, ponds and marshes. However, most breeds can be raised without swimming water. Domestic duck has a low tolerance towards salt and must therefore be supplied by fresh water.

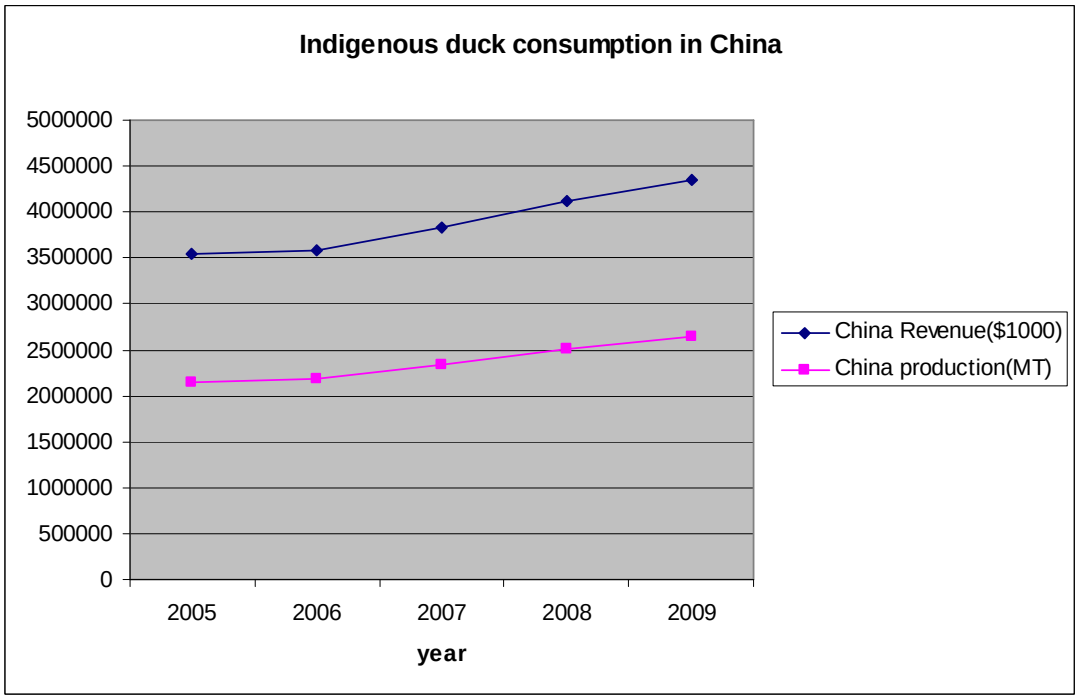


Figure I-1a: the indigenous duck consumption in China from the year 2005 to 2009. data obtained from FAO. MT: million tons. The revenue from duck production in China is indicated by blue dots.

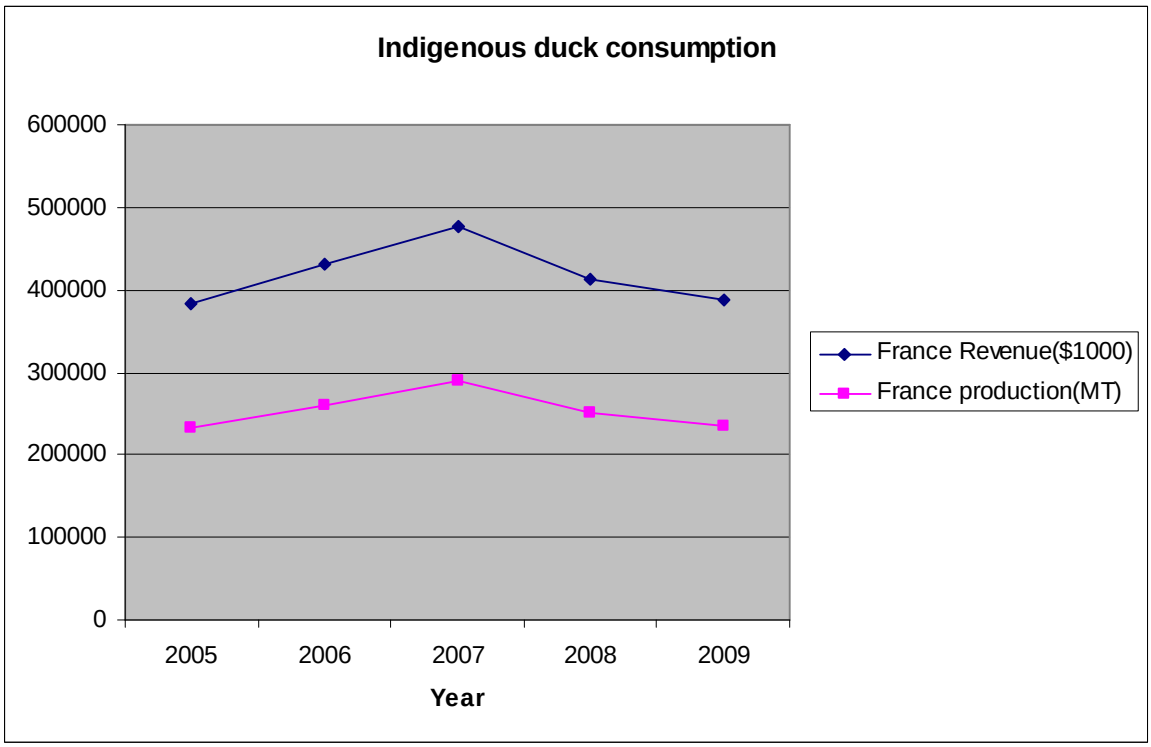


Figure I-1b: the indigenous duck consumption in France from the year 2005 to 2009 data obtained from FAO. MT: million tons; The revenue from duck production in France is indicated by blue dots, which is about 10 times less than that of China.

Ducks are also well known for their smooth temper. They are shy, nervous and only seldom aggressive to each other or to human. Most domestic ducks, especially the egg laying strains, have little instinct to brood, and as a consequence, they can lay eggs wherever they happened to be – occasional even while swimming (1991).

1.3 Duck breeding

Poultry meat represents about 33% of the global meat production: in 2007. Some 269 million tons of meat were produced worldwide, of which 88 million tons were from poultry. Chickens, turkeys and ducks are the most common sources of poultry meat (87%, 6.7% and 4% of total poultry production, respectively). However, other commercially available poultry meat include geese, pigeons, quails, pheasants, ostriches and emus (combined about 2.7% of total poultry production) (FAO's data from http://www.fao.org). In China and France, duck meat is the second most important poultry meat consumed after chicken, so duck plays an important role in agro-economics in both countries.

1.3.1 Duck breeding in China

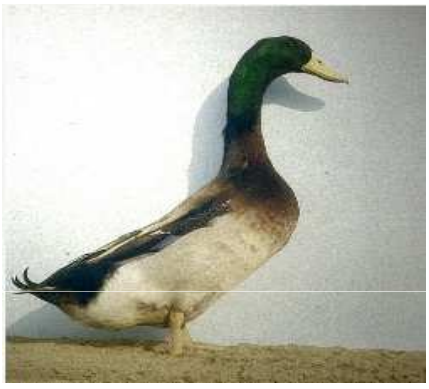
Domestication of duck in China occurred more than 3,000 years ago, not only caused by the high prevalence of wetland environment, but also because ducks have many interesting agronomic characteristics, such as a high feed conversion efficiency and growth efficiencies, good disease tolerance, and a short breeding cycle. Finally, they are easy to breed. China is by far the leading country with an annual production of about 75% of all duck slaughtered and about 66% of duck meat produced in the world (FAO's data). Duck meat consumption in China has increased in the recent years. Between 2005 and 2009, the indigenous meat consumption has increased at an average pace of 5.3% each year and by the year 2009; the

Mallard Duck (**wild type**)
(female: left, male: right)

Jinding Duck
(egg-type)

Youxian Sheldrake
(egg-type)

Liancheng White
duck
(egg-type)



Gaoyou Duck
(dual-type:
egg+broiler)

Shaoxing Duck
(egg-type)

Jianchang
Duck
(broiler)

Putian Black Duck
(egg-type)

Beijing Duck
(broiler)

Figure I-2a: Main Duck Breeds in China

production of duck meat in China was estimated to reach around 2,149,837 million tons (Figure I-1a).

In the north, duck is mainly used for the famous “roast duck” which is considered as the speciality of Beijing. For roast duck, the most widely used duck breed is the common Beijing duck (*Anas platyrhynchos domestica*). Moreover, there are some strict requirements such as the weight of the duck, which should be greater than 5kg within 100 days after hatching. Beijing ducks are mainly force-fed and used as the main breed in the north. In the south, which is the main duck breeding area in China, duck meat and eggs are frequently consumed and the popularity varies depending on different “cuisine” habits. According to some Chinese literature concerning duck breeding in China, due to differences in market demand, there are numerous domestic duck breeds in China and a survey showed that there are 27 indigenous breeds, two introduced breeds and a few breeds being recently developed. Seventy percent of the recorded breeds are distributed in southeastern China and 8 breeds (Beijing duck, Youxian Sheldrake, Liancheng white duck, Jianchang duck, Jinding duck, Shaoxing duck, Putian black duck and Gaoyou duck) have been included in the National Genetic Resource Protection program (shown in Figure I-2a). The Beijing duck is the most famous and widely used breed for broiler, because of its very fast growth rate and early development of feather and fatty tissues due to the adaption for life on water. The egg production of Shaoxing Duck and Jinding Duck is among the highest in the world, with annual production rates of 280~300 eggs of 68~70g each, though the adult body weight is about 1.3kg and 1.7kg for Shaoxing and Jinding duck respectively. Gaoyou duck is a dual-purpose breed and average adult weight is around 2.3kg for male and 2.6kg for female. Gaoyou duck is widely used to produce traditional Chinese duck meat products such as Nanjing cooked duck and Nanjing dry-cured duck which are very famous in South Asia and China. Jianchang duck is a broiler breed which is mainly kept in the Sichuan province. The

average adult weight for male is 2.4kg and for female is about 2.0kg. In addition, saline duck is frequently made from this breed; moreover, this breed is capable to produce fatty liver.

In addition to the large consumption of duck meat, there is also a large requirement for duck feather down in China. Feather down is a valuable by-product which can be used as fillers for pillows, comforters and winter clothing. Up to 2008, the production of feather down in China was of 360,000 tons in which 75% come from duck. The feather down industry provides 1.8 billion dollars for the workers and represents 55% of the world production.

Furthermore, the so-called duck-rice farming system is well-established in the south. In this system, the special fondness of ducks for mosquitos, beetle larvae, grasshoppers, snails, slugs and crustaceans is used to profit and they are used as effective pest control agents which lead to the reduction of use of pesticides. Moreover, duck feces are organic fertilizers and more ecological for rice breeding.

1.3.2 Duck breeding in France

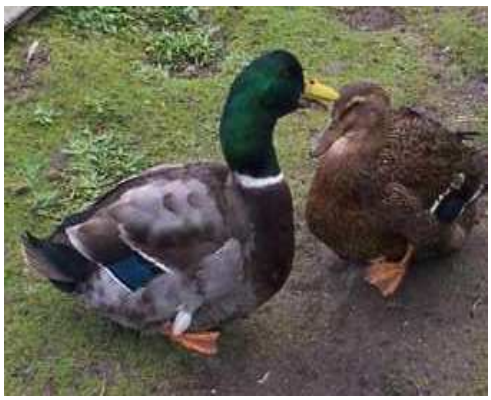
France is the second largest duck consuming country in the world. Duck breeding in France is directed mainly towards the 'foie gras' (fatty liver) industry. Various genera, species, breeds and strains of geese and ducks are bred in accordance with production requirement, but mainly two genera of ducks and their hybrids are used: the Bejing duck (*Anas platyrhynchos domestica*), the Muscovy duck (*Cairina moschata*), and the hybrid mule duck (the progeny of a Muscovy drake with a Beijing duck female). The indigenous consumption of duck from the year 2005 to 2009 in France is shown in Figure I-1b.

Mallard Duck (**wild type**)
(female: left, male: right)

Wildtype
Muscovy duck

Beijing Duck
(broiler)

Muscovy duck



Rouen duck

Mule duck
(female Beijing X Muscovy drake)

Hinny duck
(Beijng drake X female Muscovy)

Figure I-2b: Main Duck Breeds and their crosses in France. Mule ducks in France are mainly bred for Foie Gras (fatty liver) production. Hinny ducks are rarely raised in France.

About 95% of duck fatty liver production comes from force-fed male mule ducks, the remaining 5% being from the male Muscovy ducks. Recognized quality of the mule duck (rusticity, force-feeding ability, high weight and good quality of fatty liver) and the development of artificial insemination have allowed its wide-spread popularity. The French production of fatty liver was of 19,800 tons in 2008 representing 75% of the world production. The average increase of fatty liver production for the last 10 years was around 6% per year, with a very strong increase of duck liver production, whilst a long-lasting trend of reduction of goose liver production was observed. For the meat production, heavy Muscovy ducks are bred in France, the male Muscovys for cut-up pieces, and the females sold as roast ducklings. French duck meat production reached 289,792 million tons in 2007 (FAO's data) which is the most fruitful year in this decade, with around 60% coming from the fatty liver industry, the remainder from the specific duck meat industry. Unlike China and other south Asia countries, Beijing duck breeding for meat production is non-existent.

Beijing ducks have a better performance for feed efficiency and behavior traits whilst Muscovy ducks have a good force-feeding aptitude. Cross-breeding is used to produce hybrid ducks combining the merits of both parental lines for fatty liver production, resulting from heterosis effect. The progeny of a Muscovy drake crossed with a Beijing duck female are called Mule ducks, whereas the products of the reciprocal cross: a Beijing drake and a female Muscovy are named Hinny ducks (Figure I-2b). Both hybrids are usually infertile, which can be explained by the genetic distance between the parents. The female and male Mule ducks are both infertile. The male have normal testicular development and sexual activity, but do not produce spermatozoa (Snapir et al. 1998). The females do not have completely developed ovaries and any follicles; no ova are produced even though the reproductive organ exists. For the male and female Hinny duck, the situation is slightly different: the males don't produce spermatozoa and females do produce ova but cannot usually lay fertile eggs. The infertility of

both hybrid ducks could be explained by the failure of the two parental chromosomes to pair during meiosis in the hybrid germ line cells, despite very similar karyotypes (Denjean et al. 1997). This causes chromosomal sterility which differs from genetic sterility where the parental alleles lead to disharmonies of the development and sterility (Marie-Etancelin et al. 2008).

Tai and Rouvier studied the growth rate of the two species and the two hybrids in 1998, the results showed that the female Muscovy influenced the female Hinny by slowing down the growth whereas there was no impact on male Hinny duck. For the Mule duck, it seemed there was no sexual dimorphism. Comparing the same stage of growth, male Mule ducks have better performance than that of male Hinny and so does the female Mule ducks (Tai and Rouvier 1998). Finally, the mule ducks are selected in the fatty liver industry.

1.4 A scientific model for avian influenza study

Duck is not only an important agronomic species, but also a scientific model for avian influenza studies, being a natural host for avian influenza viruses. There are three types of influenza viruses: A, B, and C. Only influenza A viruses can infect birds, and wild birds are natural hosts for these viruses. Avian influenza is caused by type A viruses of the Orthomyxoviridae family. The influenza A viruses infect primarily free-living aquatic birds. Waterfowl can be infected by a very high diversity of influenza viruses and infection in wild birds is nearly always asymptomatic. The influenza A viruses are classified by their hemagglutinin (HA) and neuraminidase (NA) surface glycoproteins. All 16 HA and 9 NA subtypes have been isolated from aquatic birds, of which mallard ducks are a main reservoir (Kim et al. 2009). Type A influenza A can be further classified into two categories: low

pathogenic avian influenza (LPAI) and highly pathogenic avian influenza (HPAI). LPAI can infect human and birds but only provoking mild symptoms, whereas HPAI cause severe illness and high mortality in poultry. However, LPAI have the potential to evolve into HPAI. Among the subtypes isolated, H5 and H7 are of particular concern because they can become highly pathogenic, causing systemic illness and death in both avian and mammalian species, including human (Swayne and Suarez 2000).

The prevalence of avian influenza is mainly due to ducks migration during spring and autumn, influenza viruses having thus the potential of being transmitted along the migration route to the domestic duck populations. Thereafter, infected domestic ducks are likely to maintain the virus locally, which may then spread to other species (Kim et al. 2009). The virus replicates in the cells lining the intestinal tract and is excreted at high titres in the feces of the infected ducks that do not show clinical signs of disease and scarcely produce detectable serum antibodies. All human influenza pandemics can be traced back to viruses that originated in ducks (MacDonald et al. 2007). Most of HPAI viruses are 100% lethal to chickens and gallinaceous poultry, they often cause asymptomatic infection in some species of domestic and wild ducks, which can still be a serious problem for the duck industry and can lead to the slaughtering of all animals the farms in which outbreaks occur (Kida et al. 1980; Songserm et al. 2006).

The best known HPAI, the H5N1 HPAI virus, which emerged in Asia in 1996, is unique among the highly pathogenic avian influenza (HPAI) viruses in that it has continued to circulate in avian species for more than a decade and has spread to more than 60 countries in Eurasia. Moreover, it is evident that some strains of H5N1 HPAI can also infect human and cause lethality. Around 60% of human individuals infected by H5N1 HPAI have died from it and furthermore, there is a possibility that H5N1 may mutate into even more highly

pathogenic strains, capable of efficient human-to-human transmission. According to the FAO Avian Influenza Disease Emergency Situation Update, H5N1 HPAI pathogenicity is continuing to rise gradually in endemic areas although the avian influenza disease situation in farmed birds is being held in check by vaccination. Eleven outbreaks of H5N1 HPAI were reported worldwide in June 2008 in five countries (China, Egypt, Indonesia, Pakistan and Vietnam) compared to 65 outbreaks in June 2006 and 55 in June 2007. Each outbreak leads to tens of thousands of poultry slaughtered and incinerated and consequently leads to the loss of farmers and/or poultry breeding enterprises.

Due to the high lethality and virulence of HPAI A (H5N1), its endemic presence, its increasingly large host reservoir, and its significant ongoing mutations, the H5N1 virus is the world's largest current pandemic threat and billions of dollars are being spent on the study of H5N1. Genetic evidence shows that H5N1 HPAI viruses originated from a H5 LPAI virus from a wild mallard or another migratory wild bird in northern Japan (Duan et al. 2007; Okazaki et al. 2000). In the first place, H5N1 HPAI killed ducks as well, but maybe due to the specificities of the immune response in ducks or to the rapid adaption of viruses, H5N1 HPAI became less pathogenic to duck (Hulse-Post et al. 2007; Hulse-Post et al. 2005). The first case of human lethality caused by H5N1 HPAI virus was reported in Hong Kong in 1997, which was a direct bird to human transmission (Peiris et al. 2007). To control the avian influenza pandemic, apart for producing vaccines, many efforts also have been made on understanding the genetic basis of the duck's resistance to those viruses.

A recent study made by Magor *et al* suggests that the influenza virus sensor, retinoic acid inducible gene I (*RIG-I*), is present in ducks and plays an important role in clearing an influenza infection (Barber et al. 2010). *RIG-I* encodes a *DExD/H* box RNA helicase that contains a caspase recruitment domain as an essential regulator for dsRNA-induced signaling

which transmits downstream signals and thus results in the activation of transcription factor *NF-κB* and *IRF-3*. Subsequent gene activation by these factors induces antiviral functions such as type I interferon production (*IFN*) (Yoneyama et al. 2004). *RIG-I* is pivotal in detection and eradication of replicating RNA virus genomes, as showed by the fact that dsRNA viruses were more virulent and active in *RIG-I*-deficient mice than in controls (Kato et al. 2006). Duck *RIG-I* has features in common with mammalian *RIG-I*; duck *RIG-I* is 933 amino acids and in human is 925 (Barber et al. 2010). Domain prediction also shows that duck *RIG-I* has an N-terimal caspase recruitment doman, a helicase domain and a *DExD/H* box helicase domain, consistent with the mammalian structure (Takahasi et al. 2008; Yoneyama et al. 2004). A striking finding is that *RIG-I* is apparently absent in chicken which may explain why the chicken has higher mortality than duck after avian influenza infection. The presence of *RIG-I* in ducks demonstrates that an early antiviral response may contribute to survival of lethal avian influenza infection.

These may be the tip of iceberg in the signaling pathway induced by avian influenza viruses, more efforts are still needed to be made to unravel the immunogenetics of the relevant response.

1.5 The rationale for duck genomics

As public concern for animal welfare become more and more critical, duck breeding is now facing a constricting regulation. For instance, in Europe, the permanent comity of the European convention on breeding animal welfare is asking for evolutions in the duck's housing systems during the force-feeding phase and more specifically a change from individual to collective housing by 2016. On very a short term, the industry must therefore

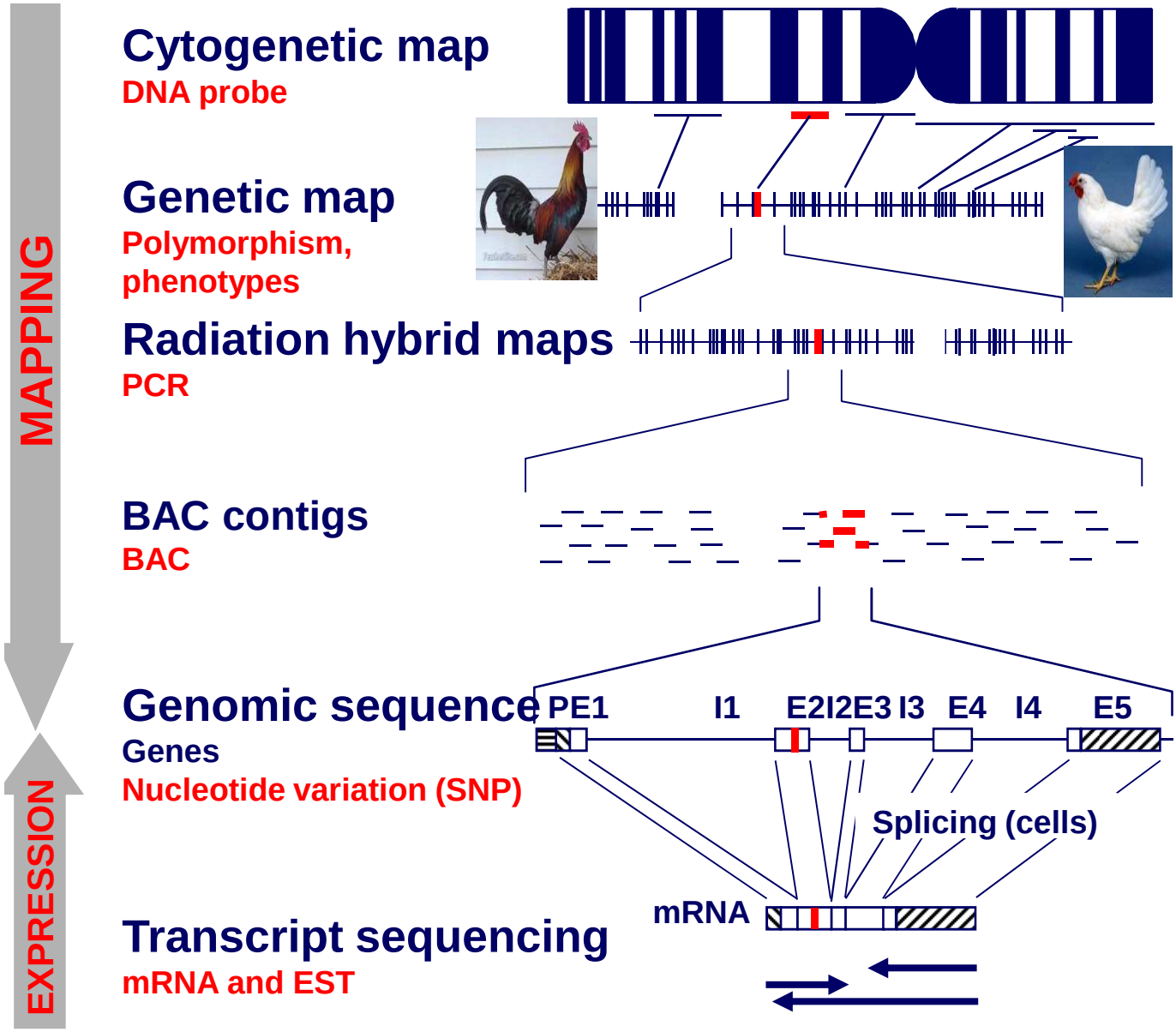


Figure I-3: Genome maps. Each genome map type provides different information. Cytogenetic mapping, by hybridizing directly DNA probe(s) onto chromosomes, provides the chromosome identification. Genetic mapping, reflecting recombination frequencies in meiosis between markers, requires polymorphic markers and can integrate phenotypic data. Radiation hybrid mapping is based on physical chromosome breakage by radiations. It uses PCR for mapping markers and thus can allow mapping of any STS markers (microsatellite, EST…). BAC contig mapping is a physical map providing the information on the ordering of BAC clones, which can be sequenced, along the chromosomes. All those maps are invaluable ressources for the obtention of a complete genome sequence. EST and mRNA are indispensable for gene annotation.

adopt new housing systems, to which the duck stocks must be adapted. This necessary adaptation may imply a relaxed selection pressure on production traits. The detection of genome regions influencing the duck's behaviors during the breeding and force feeding phases could help adapting stocks to new breeding systems. Moreover, the detection of QTL controlling production traits could enable to limit the loss of genetic gain in the selected breeds.

Due to more and more constricting regulation in the animal breeding industry, genetics could be a powerful tool to handle the challenges. But, before the year 2000 hardly any molecular genetic markers were known for duck. So as to initiate molecular studies on the genetic variability, such as QTL detection in duck, the Laboratoire de Génétique Cellulaire and the Station d'Amélioration Génétique des Animaux at INRA Toulouse developed respectively specific duck microsatellites and a resource family for QTL detection of the traits related to fatty liver production. In China, the state-key laboratory for agro-biotechnology in China Agricultural University collaborated with the most famous duck breeding company Golden Star Duck Production (Beijing) and established also a resource family to detect QTL traits related to growth. More and more efforts are dedicated to duck genomics world-wide, and the duck genome has been sequenced by using Illumina GenomeAnalyser *II* sequencing machines at the Beijing Genomics Institute (BGI) by Huang *et al* (Huang et al, in prep). In addition to current efforts in China and France, the sequencing of EST libraries from immune tissues is undergoing at the Roslin Institute, UK and the University of Alberta, Canada. The production of SNP by sequencing northern and southern European mallard duck samples has been published by Kraus et al in Netherland (Kraus et al. 2011).

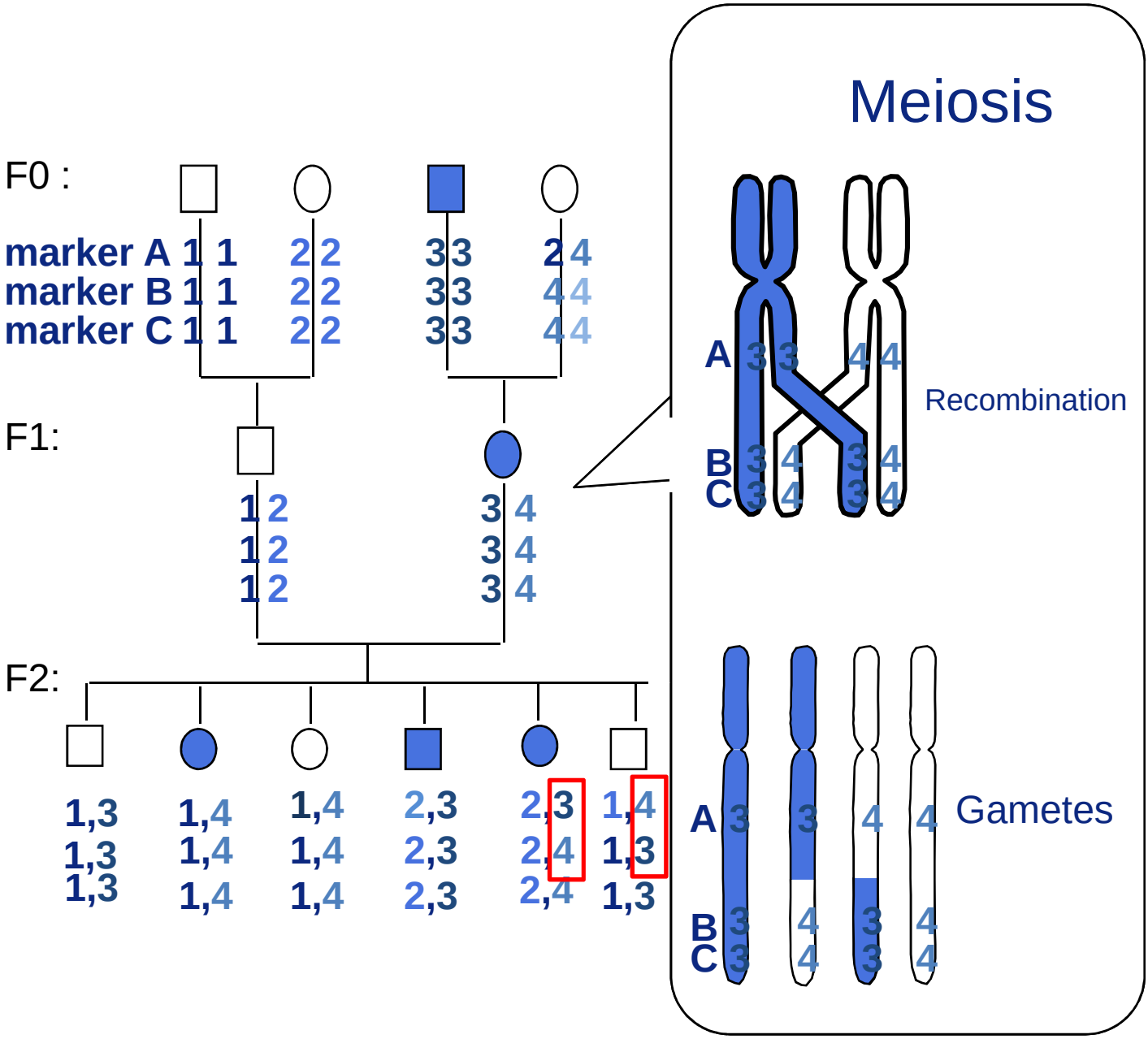


Figure I-4: Principle of genetic mapping. Genetic maps are constructed by linkage analysis between markers segregating in families. This schema is a brief view about crossing over and recombination events happening in meiosis. Numbers 1,2,3 and 4 stand for different alleles inmarkers A, B and C. The phenotype highlighted in blue is dominant An example of a recombination events between markers A and B in the F1 female (right) gives recombinant gametes. Recombination events can be detected in the F2 off spring by genotyping the whole family with the markers. The red boxes in F2 show the recombination events detected. The frequency of recombination events is a function of the distance. The blus phenotype segregates with allele 3 of marker 1, excluding the distal part of the chromosome for its localization.

2. Genome mapping and sequencing

A genome map allows navigating around a genome and consists of a set of molecular marker landmarks. The different genome mapping techniques have a range of resolutions and one or another will be used according to the question to be solved and to the available data and resources (Figure I-3). Genome maps are divided into two main categories: the genetic maps and the physical maps. Genetic maps are used to order loci along chromosomes on the basis of the frequency of meiotic recombination events. The observation of such recombination events is made by observing the segregation of alleles at different loci from parents to offspring. These alleles can either be phenotypic differences due to polymorphism at singular loci (shape of organs, color…) or the visualisation of allelic differences at the molecular level, usually on the DNA sequence itself (microsatellite, SNP markers…). The highest the frequency of co-segregation of alleles at two tested loci, the closest they are considered to be on the chromosome. Two loci segregating independently are either on different chromosomes or far from each other on the same chromosome (Figure I-4). The distances on a genetic map are thus estimations and the link to physical distances is dependent on local recombination rates, which can vary along chromosomes. The physical map is a representation of the chromosome providing the physical distance at the DNA level between markers on the chromosomes. The main categories of physical maps are the cytogenetic map in which the chromosomes are visualized under a microscope, the Bacterial Artificial Chromosmes (BAC) contig maps, the radiation hybrid (RH) maps and the ultimate map represented by the genome sequence.

To construct different maps, genetic markers are prerequisite components which tell apart cells, individuals, populations or species. There are several types of markers used in the

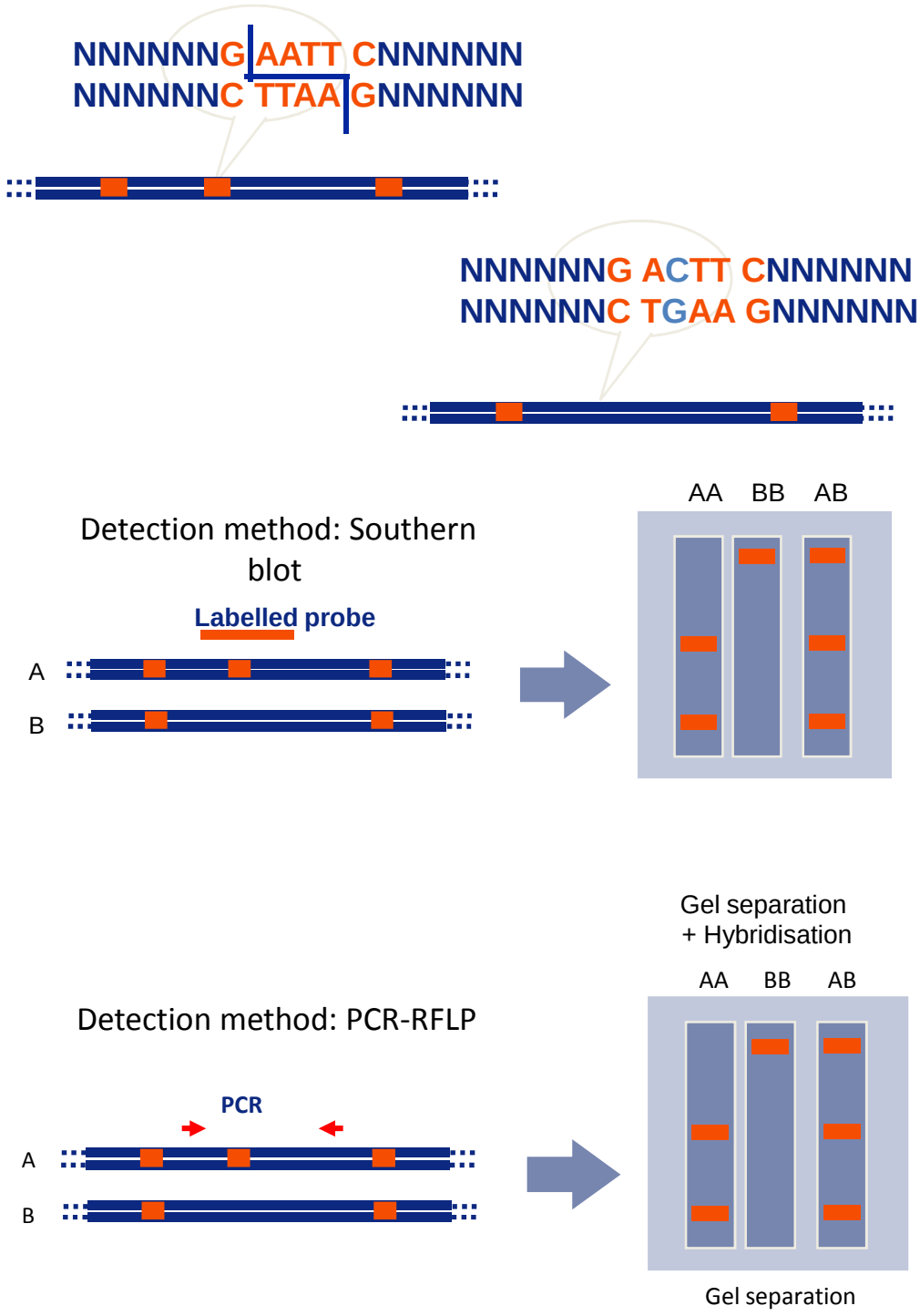


Figure I-5: Restriction Fragment Length Polymorphism (RFLP). They were formerly used as DNA markers for linkage analysis. At a given locus, one chromosome carries three "GAATTC" restriction sites whereas another bears a mutation in the second restriction site. Before the invention of PCR, Southern blot was used to verify the presence or absence of the restriction site. Examples of homozygote (AA or BB) and heterozygote (AB) individuals are given.

genome mapping; the most common types are restriction fragment length polymorphism (RFLP), simple sequence length polymorphism (SSLP) and single nucleotide polymorphism (SNP).

2.1 Genetic markers

The first genetic maps were constructed at the beginning of the 20th century by using phenotypic data (Hutt 1933) and the first genetic map published for livestock species was a map for chicken composed of 18 markers in 5 linkage groups by Hutt (Hutt 1936). This so-called classical map was updated regularly and the least versions was composed mainly of phenotypic markers and mutations observed in chicken lines and of blood group loci by Bitgood *et al* (Bitgood and Somes 1993). After these first efforts, more extended genetic maps were developed in several livestock species by using molecular markers revealing polymorphism directly at the DNA level.

RFLP (Restriction Fragment Length Poluymorphism) are the first type of DNA markers to have been largely used for genetic mapping. Restriction fragments are produced when a DNA molecule is digested by a restriction endonuclease, a type of enzyme which cuts DNA at a defined sequence, for instance GAATTC (Figure I-5). Many different restriction endonucleases exist and restriction fragments produced at a defined locus can be detected by Southern blot. Any change in the target sequence will affect the enzyme's cutting activity and thus DNA sequence polymorphism can be detected by the change in the size of restriction fragments between two alleles. The major drawbacks of RFLP are that it requires a mutation to be within a restriction enzyme target sequence, that only two alleles per locus can be detected and also that the detection technique, although improved by the use of PCR-RFLP, is difficult to use on a large scale with many different markers.

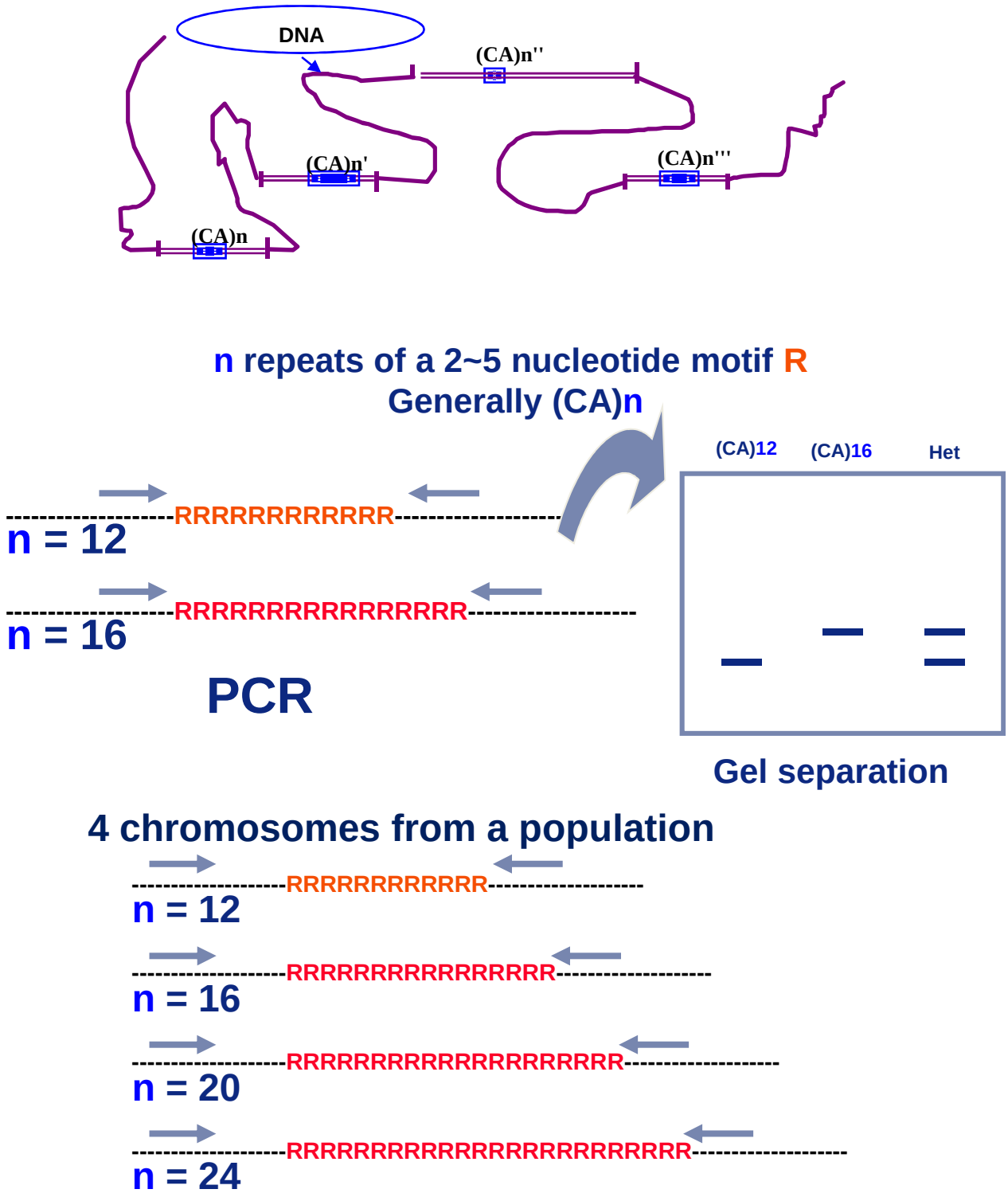


Figure I-6: Microsatellite markers. Several thousands of microsatellite markers are dispersed all along the genome. A given microsatellite can be amplified with locus-specific primers. One advantage of microsatellites is the high number of alleles that can be observed at one locus.

SSLP (Simple Sequence Length Polymorphism) are arrays of tandem repeat sequences displaying length variations, the different alleles containing different numbers of a repeat unit. SSLP can be multiallelic since each SSLP can have a number of different length variants. There are two types of SSLP: minisatellites also known as Variable Number of Tandem Repeats (VNTR) in which the repeat unit is a few tens of nucleotides in length and microsatellites, which were the most widely used markers for the construction of genetic maps in the past two decades, containing repeat units of a few bases in length. The reasons why microsatellites are more popular than minisatellites are that microsatellite are more abundant and evenly spread out along the chromosomes while minisatellites are more prone to be found near the ends of the chromosomes and that microsatellites can be easily genotyped by PCR (Figure I-6).

Single Nucleotide Polymorphisms (SNP) are now the most commonly used genetic markers. As suggested by the name, an SNP marker is a single base change in the DNA sequence, with a usual alterative out of two possible nucleotides at a given position. Several million of these single nucleotide substitutions are known for mammalian species such as human or mouse. On average, when comparing two chromosomes picked at random in a population, one SNP every kb can be found in human and one every 200 pb in chicken (Wong et al. 2004). For such a base position with two alternatives in genomic DNA to be considered as an SNP (versus a mutation), it is considered that the least frequent allele should have a frequency of 1% or greater (Vignal et al. 2002). In the genome, the density of SNP is higher than that of microsatellites and genotyping techniques allowing the simultaneous analysis of several hundred alleles have been developed. Thus, by using SNP as genetic markers, high density genetic maps can be built facilitating genetic approaches.

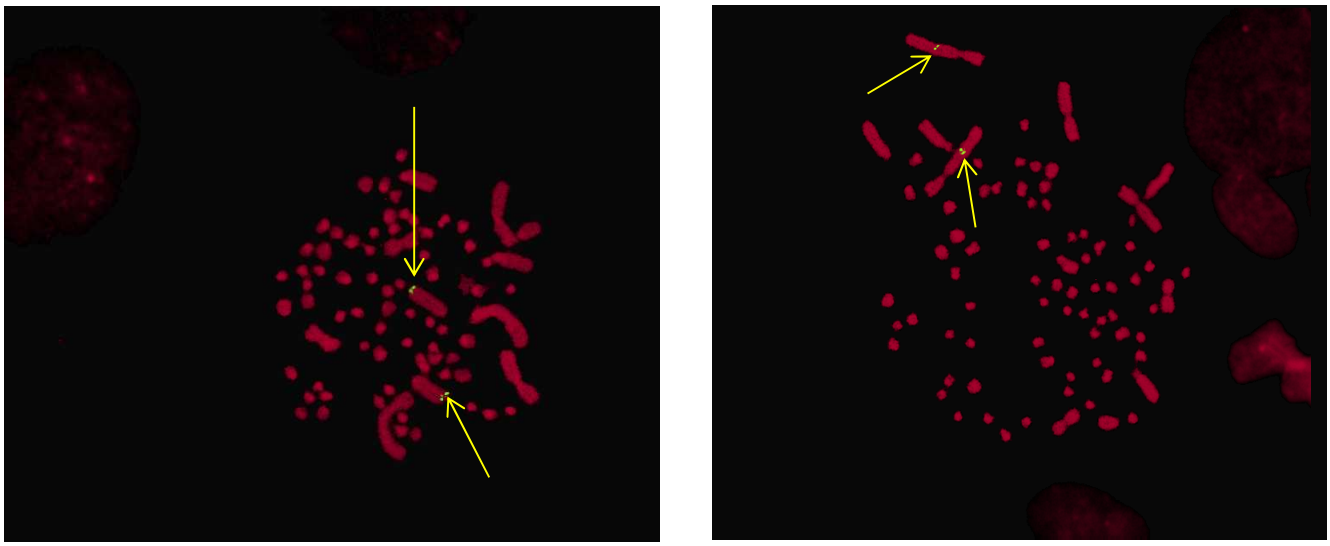
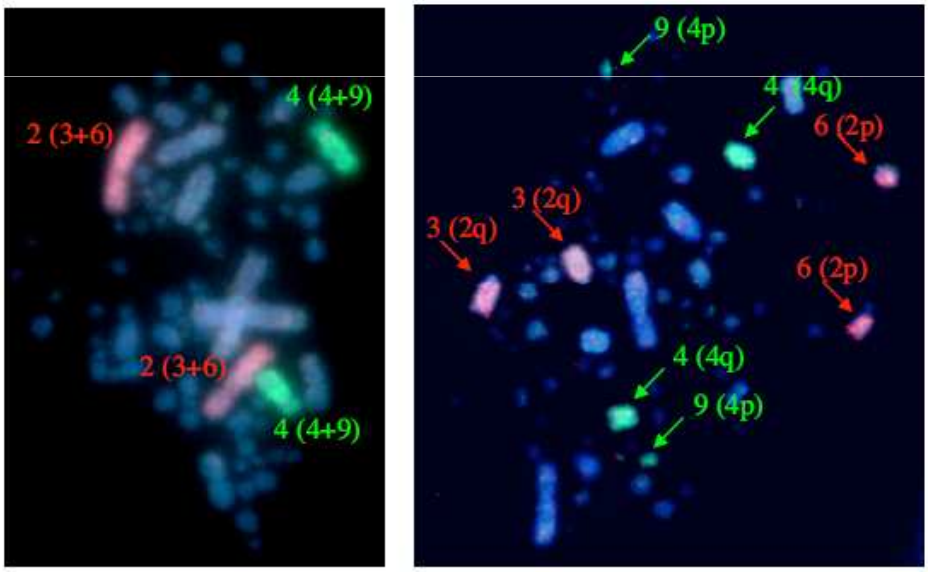


Figure I-7a: Fluorescent in situ Hybridization (FISH). Chromosomes are arrested at metaphase. The green signal showed above are probes hybridized to chromosomes. Yellow arrows indicated the signal



Chromosome paints for chicken chromosomes GGA2 (red) and GGA4 (green). a) On a chicken metaphase (chromosome numbers are labeled with turkey (MGA) orthologues in brackets). b) On a turkey metaphase (chromosome numbers are labeled with arrows and chicken (GGA) orthologues are indicated in brackets).

Figure I-7b: an example for interspecific painting between chicken and turkey. (Cited from D. Griffin et al., **Whole genome comparative studies between chicken and turkey and their implications for avian genome evolution** BMC Genomics 2008)

2.2 Cytogenetic, BAC contig and genetic maps

Cytogenetic techniques allow the visualisation of chromosomes condensed at the metaphase stage under a microscope. The routine analyses involved in cytogenetic mapping are mainly banding techniques, with alternate bands obtained by staining techniques, and fluorescent in situ hybridization (FISH), allowing the positioning of DNA fragments on chromosomes. A FISH example was shown in Figure I-7a. Banding technique provides an over view of chromosome size and band patterns allow to distinguish chromosomes of similar size and gives an idea of the structure of the chromatin, e.g. in G-banding, the deeply stained fragments are heterochromatin which are more compact while the less stained are euchromatin. FISH allows the mapping of DNA fragments of a minimum size of 10-20 kb. Its resolution is low when compared to other mapping techniques, but it is the only one allowing the assignment of loci to chromosomes. Thus other map types can be linked to chromosomes via FISH (Figure I-3). Inter-specific mapping, allowing the rapid discovery of evolutionary breakpoints, mainly translocations and inversions, can be done by standard FISH or by chromosome painting, in which a probe for an entire chromosome in one species is hybridized to a metaphase of another species (Figure I-7b).

Genetic maps are based on the calculation of recombination frequencies between markers by linkage analysis. During meiosis, homologous chromosomes pair and the sister chromatids are exchanged at points of crossing-over as demonstrated in Figure I-4. The recombined chromosomes subsequently segregate into the gametes. The recombination rate between two markers will increase with distance and the distances thus measured in genetic map are expressed in centimorgans (cM). One Morgan is the distance between two markers at which one recombination event will be observed statistically in each meiosis. Recombination rates vary along chromosomes and high recombination rates within short physical intervals

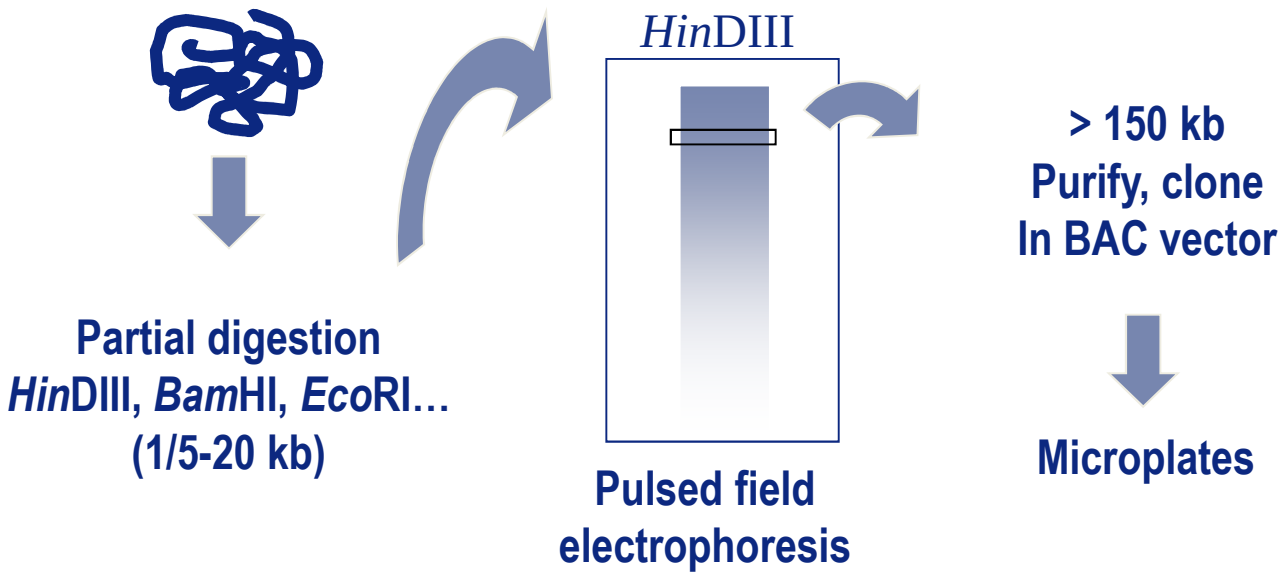


Figure I-8a: Construction of a BAC library. Partial restriction digests of genomic DNA are separated by pulse-field gel electrophoresis. Large fragments are isolated and inserted in a BAC cloning vector. Individual clones are kept in microplates. A typical library will cover at least 5 to 10 genome-equivalents.

can be detected and are named recombination hotspots. Therefore, genetic maps reflect the situation of the recombination rate during meiosis which sometimes is not always directely correlated to physical distances. Genetic maps are especially useful, as they allow determining the position of a gene even though nothing is known about it except for its phenotypic effect (Dear 2001). Therefore, genetic maps are invaluable for the localization of genes responsible for specific traits, such as genetic diseases or traits of agronomic importance. However, the construction of genetic maps require one or several genetic mapping populations including at least three generations in which the segregation of genetic marker can be observed and it also requires polymorphic markers: RFLP; microsatellites or SNP, allowing the distinction of the parental origin of alleles at the loci studied.

The BAC contig maps are physical maps in which more than 100-500 thousand BAC clones (Figure I-8a), containing each over 100 kb of DNA are ordered along the genome. BAC clones are subjected to restriction digestion, to produce fingerprints after separation by gel electrophoresis and the overlapping clones share a subset of fragments, by which the order can be inferred (Figure I-8b). BAC fingerprint mapping has been successfully applied for the human genome by Gregory *et al* (Gregory et al. 1996) and for the chicken genome by Wallis *et al* (Wallis et al. 2004).

2.3 Genome maps using somatic cell radiation hybrids: a history

2.3.1 Radiation hybrid map

Whole genome radiation hybrid mapping became a mainstream mapping technique for high resolution gene mapping in mammals in the beginning of the 90's (Cox et al. 1990; Gyapay et al. 1996). RH mapping is used to complement linkage and other physical maps by

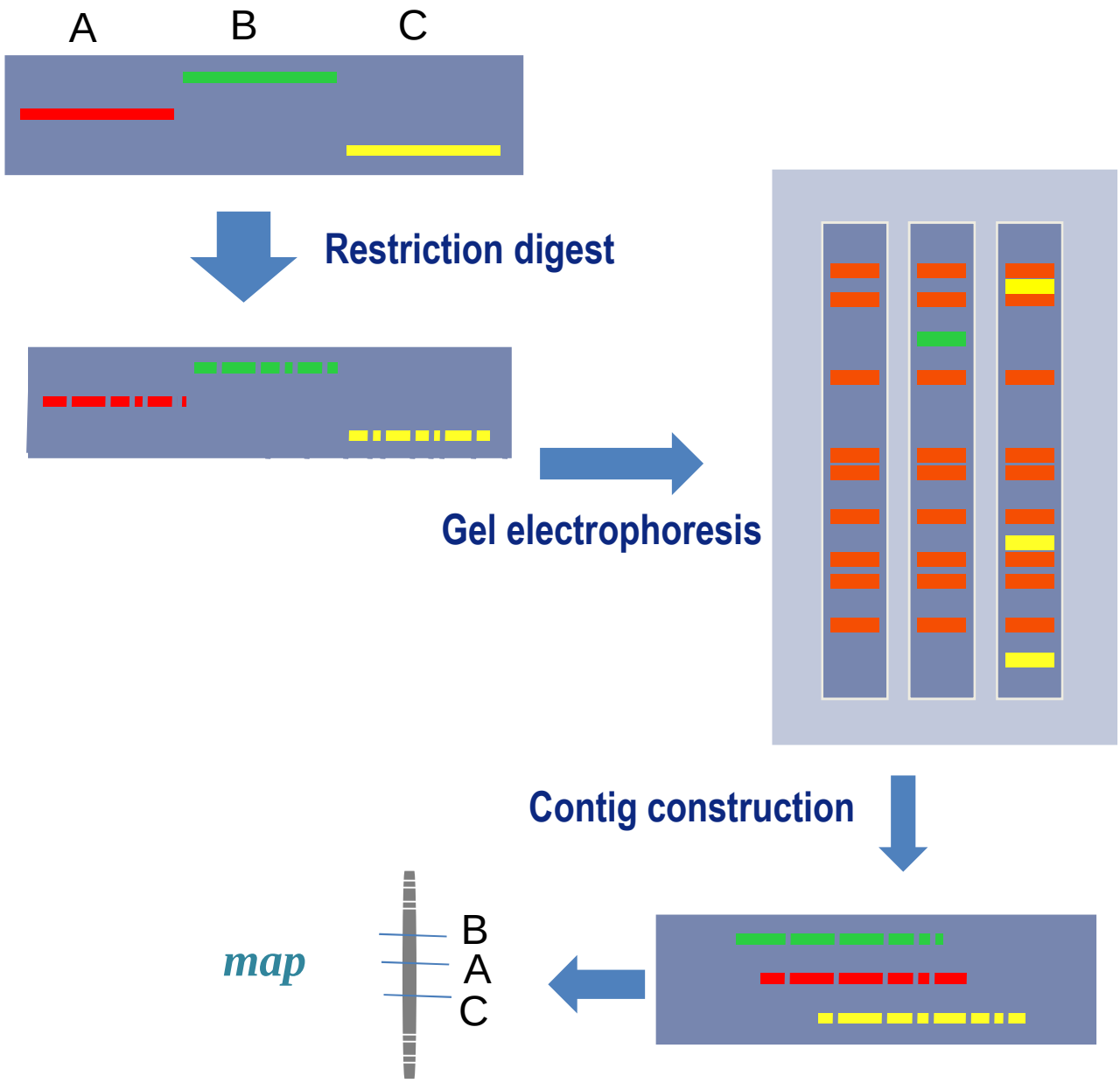


Figure I-8b: The principle of BAC contig construction. Three clones used as a demonstration example. After restriction enzyme(s) digestion, the small BAC DNA fragments are separated by electrophoresis to generate fingerprints. Overlapping BAC clones share a high number of restriction fragments. Fingerprint comparison of all BAC clones from a library allows contig building over the entire genome.

providing higher resolution ordering than usual genetic and cytogenetic maps and thus is an aid for the assembly of whole genome sequence contigs and scaffolds (Hitte et al. 2005). RH mapping also holds great utility for generating comparative maps in species for which the development of crosses for genetic mapping is logistically problematic (Page and Murphy 2008).

2.3.2 History

RH mapping can be traced back to the 1970s, when the concept of "radiation hybrid (RH)" cells was first proposed by Goss and Harris (Goss and Harris 1975), as an extension of the somatic hybrid cell mapping technique. Somatic hybrid cells were made by the fusion of two parental cells by treatment with inactivated Sendai Viruses, with lysolecithin or with polyethylene glycol, agents provoking the fusion of cell membranes. The fusion between two parental cells of different species of origin gives rise to binucleate heteokaryons, which generate mononucleated or hybrid daughter cells after the following mitosis. In the 1970s, most somatic hybrids were produced between human and mouse or human and hamster cells. In these hybrids, the unilateral loss or segregation of human chromosomes was observed so that they could be used to map genes (Ruddle 1973). The use of mutant recipient cell lines, deficient in enzymes involved in the metabolism of nucleotides, such as hypoxanthine-guanine phosphoribosyltransferase (*HPRT*) or thymidine kinase (*TK*) allowed the selection of hybrids by cultivating the fusion partners in a selection medium containing hypoxanthine, aminopterin and thymidine (HAT). *HPRT* deficient cells cannot incorporate hypoxanthine whereas *TK* deficient cell cannot metabolize thymidine. If *de novo* synthesis of purines and pyrimidines is blocked by the antimetabolite aminopterin, cells become dependent for survival on exogenous hypoxanthine and thymidine. *HPRT* or *TK* deficient recipient cell are

thus conditional lethal, hence, the fusion of *HPRT* or *TK*-deficient recipient cell with normal cells expressing functional *HPRT* or *TK* yields hybrids whose enzyme deficiency is complemented and thus that can survive in the HAT medium. The donor cells which do not fuse are eliminated due to low growth rate. In the somatic hybrid cell system, no breakage or a very low breakage of donor cell chromosomes is observed and the technique was mainly used for assigning genes to chromosomes and eventually to large chromosome regions.

To allow for fine mapping of genes, a new method called irradiation and fusion gene transfer (IFGT), was proposed by Goss and Harris. The main difference when compared to the somatic hybrid cell approach is that donor cells are subjected to a large dose of ionizing radiation which breaks their chromosomes and is was lethal for the cells. Irradiated donor cells are then fused with *HPRT* or *TK* deficient recipient cell and hybrids selected on HAT medium. As the PCR technique was not yet available in 1975, Goss and Harris used as markers 4 X-linked enzymes including the selective marker *HPRT* and were able to establish the order of these four markers on the long arm of HSAX and to demonstrate retention of non-selected chromosome fragments. They also derived mathematical approaches for constructing maps based on the co-retention frequencies of markers.

However, due to the lack of markers to map and of high throughput genotyping techniques, the IFGT method was not widely used in the following years. A renewed interest in IFGT was prompted by Cox and coworkers (Cox et al. 1990), who modified the original approach by using a rodent-human somatic cell hybrid as donor cell instead of a diploid human cell. In his approach, the donor cell was monochromosomal rodent-human somatic cell hybrid containing HSA19 and a map containing 14 DNA probes spanning 20Mb was obtained and confirmed by pulse-field gel electrophoresis (PFGE). The work done by Cox and co-workers demonstrated the effectiveness of RH mapping for constructing high-resolution,

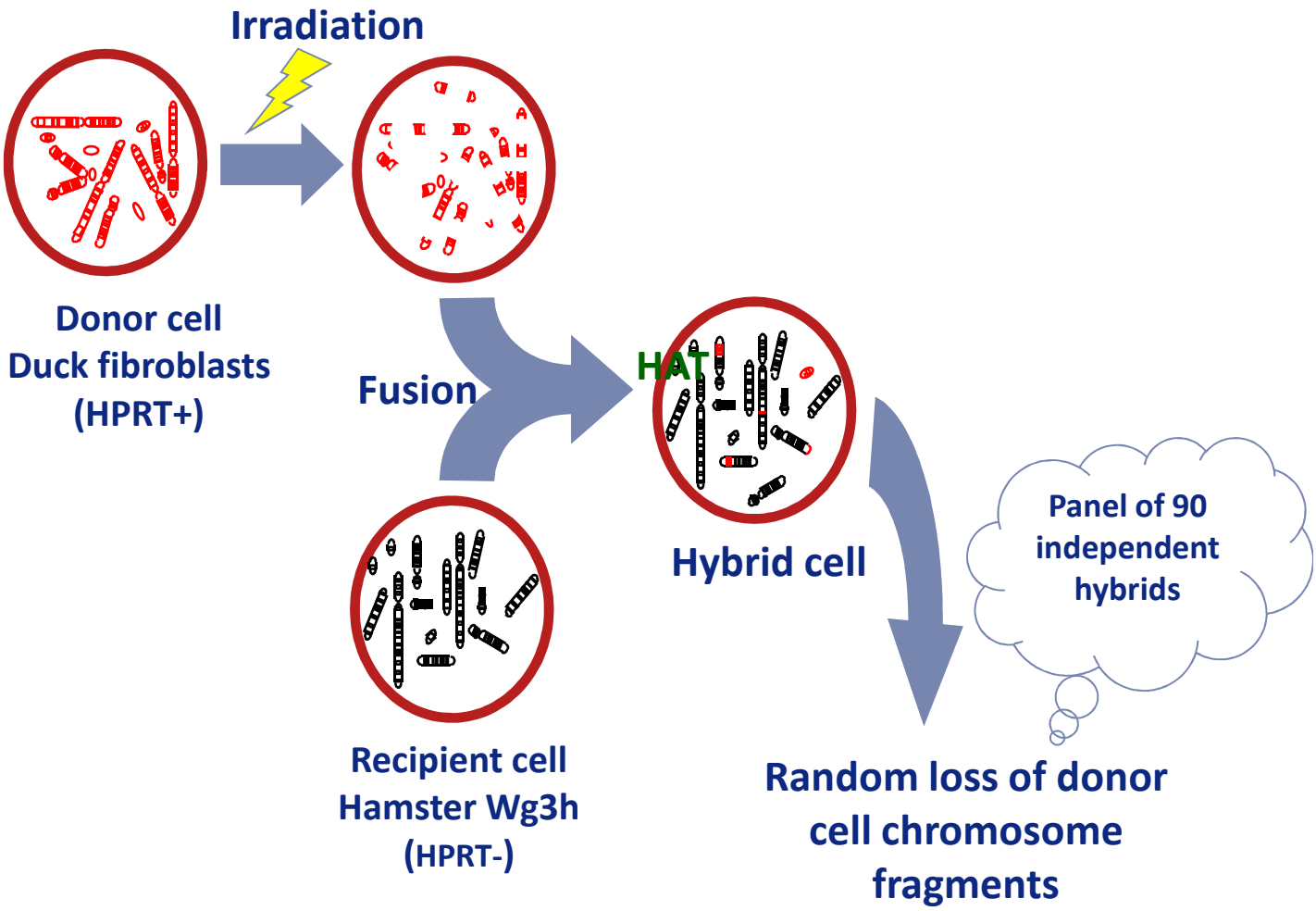


Figure I-9: principle of RH panel construction HPRT- deficient recipient cells are fused with donor cells, whose chromosomes are broken by irradiation. After fusion, all the cells are cultured in HAT selective medium and chromosome fragments of the donor cells are randomly lost. Only hybrid cells which bear the *HPRT gene* from donor cells can grow and propagate in HAT medium, loosing donor cell chromosome fragments at random. A panel comprising of 90 to 100 hybrids is used for RH mapping.

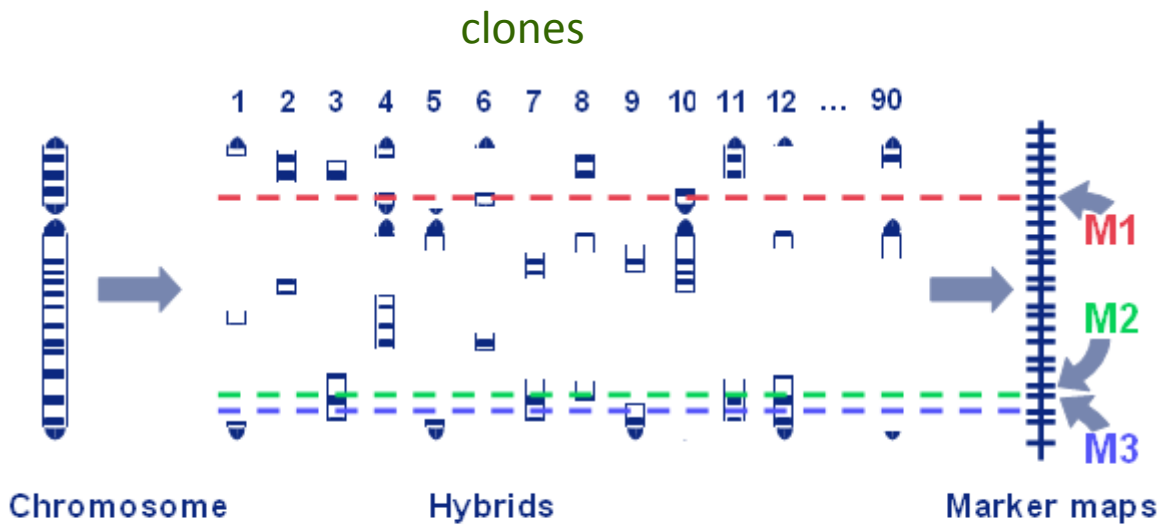
contiguous maps of a mammalian genome. An important feature of this study was the retention of human chromosome fragments without selection, allowing an equal sampling of all regions of the chromosome studied. The advantage of the protocol was that all the human fragments retained in the radiation hybrids were derived from a known human chromosome, thus simplifying the mapping process and map building. However, the major drawback was that approximately 100 hybrids were required for building a map for each chromosome, and thus for human whole genome would require thousands of hybrids, making the process very expensive and laborious.

In this context, Walter and his colleague proposed a new method which is widely used nowadays for RH mapping (Walter et al. 1994). They reverted to Goss and Harris' original protocol and used a diploid human genome fibroblast as donor cell. Forty-four hybrids were obtained and used to make a map of HSA14 containing 40 ordered markers, from which they suggested that the construction of a high-resolution map of the whole human genome was feasible with a single panel using 100 – 200 hybrids (Walter et al. 1994). The principle for developing a RH panel is described in Figure I-9.

2.4 Radiation Hybrid (RH) mapping

2.4.1 Principle

The principle of RH mapping uses the fact that after irradiation of the donor cells, the chromosomes are broken and randomly lost during cell culture (Figure I-10). After cell culture, on average 20 % of the donor cell chromosome fragments are retained in the hybrids and the probability for the simultaneous rescue of two markers will increase proportionally



PCR Results

Hybr.	1	2	3	4	5	6	7	8	9	10	11	12	...	90
M1	-	-	-	+	-	+	-	-	-	+	-	-		-
M2	-	-	+	-	-	-	+	+	-	-	+	-		-
M3	-	-	+	-	-	-	+	-	+	-	+	-		-

Figure I-10: principle of RH mapping. The example is given for one chromosome. Only fragments of the chromosome are present in each of the 90 clones representing the RH mapping panel. Each clone has randomly retained different fragments. Markers M2 and M3 are close together and tend therefore to be retained in the same chromosome fragments with only occasional breakage. Contrariwise, M1 appears independent, never being present at the same time as M2 and M3 and is either on a different chromosome or on the same chromosome than M2 and M3, but at a long distance. However, the fact that all three markers belong to the same chromosome can be deduced from the mapping of additional markers, allowing the constitution of a linkage group (right).

with their proximity on the chromosome, as the chances of having a radiation-induced break between them will decrease. If two markers tend to be retained together in many hybrid cell clones, the number of breaks is low and thus the distance will be short.

RH mapping is performed by genotyping markers on the hybrids, usually by PCR, to test for the presence or absence of the corresponding chromosome fragments (Figure I-10). The key observation behind RH mapping is that if two markers are close to each other on a chromosome, then the probability of a break occurring between them during the irradiation is smaller than if they are far apart. If no break occurs between two markers located on the same chromosome, they will be on the same fragment and the markers will either be both present (retained) or both absent (non-retained) in the hybrid cell lines. For a pair of markers, the closer they are located on the same chromosome, the higher the co-retention. A break between markers is observed when one is positive (presence of the chromosome fragment) and the other negative (absence) in a hybrid cell clone. An RH map is built by genotyping a series of markers and then calculating the relative likelihood of the proposed order of loci and/or the distances between them.

2.4.2 Published RH panels and maps

Two decades have passed since Walter et al reported the method to construct whole genome radiation hybrids. RH panels and maps are now available for many mammals including human (Gyapay et al. 1996; Olivier et al. 2001; Stewart et al. 1997), macaque (Murphy et al. 2001), mouse (Avner et al. 2001; McCarthy et al. 1997; Schmitt et al. 1996; Van Etten et al. 1999), rat (McCarthy et al. 2000; Watanabe et al. 1999), bovine (Itoh et al. 2005; Marques et al. 2007; Prasad et al. 2007; Rexroad et al. 2000; Williams et al. 2002; Womack et al. 1997), swine(Hamasima et al. 2003; Yerle et al. 1998), dog (Vignaux et al. 1999), cat (Murphy et al. 1999), horse (Chowdhary et al. 2002; Kiguwa et al. 2000),sheep

(Laurent et al. 2007; Wu et al. 2007), river buffalo (Ianella et al. 2008), but also in non-mammalian species like chicken (Douaud et al. 2008; Jennen et al. 2004; Leroux et al. 2005; Morisson et al. 2007; Morisson et al. 2004; Morisson et al. 2002; Morisson et al. 2005; Pitel et al. 2004; Rabie et al. 2004), zebrafish (Geisler et al. 1999; Hukriede et al. 2001; Hukriede et al. 1999; Kwok et al. 1999), medaka fish(Su et al. 2007), sea bass(Guyon et al.) and gilthead sea bream(Sarropoulou et al. 2007; Senger et al. 2006).

Several RH panels can be available for the same species such as for human, bovine and swine; some differences exist among these panels, such as the dose of irradiation used to break the donor cell chromosomes. The irradiation step has two functions: first, a lethal dose is necessary to kill the donor cell and to ensure that any survival cell is true hybrid and a 1500 rad dose is largely sufficient to kill most cell types; second, the irradiation causes double strand breaks in DNA and shatter the chromosomes in the cell: the larger the irradiation dose, the higher the number of breakages in the chromosomes (Walter and Goodfellow 1993). Siden *et al* using human chromosome Xq27-28 region as a model, found that 40% of the hybrids generated at 5,000 rads or less were found to have retained fragments in the range of 3-30 Mb, 10% retained whole chromosome arms, and the remaining 50% retained fragments of less than 2-3 Mb (Siden et al. 1992). The proportion of fragments of 3 Mb or larger decreased rapidly at higher irradiation doses and was very low (less than 6%) in hybrids generated at 25,000rads (Siden et al. 1992).

So, according to the purpose for which the panel is designed, different irradiation doses can be adopted. Radiation hybrids for whole chromosome mapping are generated with low doses (< 10,000rad), whereas for local high resolution mapping, such as for positional cloning experiments, higher doses (> 10,000 rad) will be used.

2.4.3 Radiation hybrids are unstable

Radiation hybrid cell lines are unstable: donor cell chromosomal fragments tend to be eliminated during the cell divisions. Karere and his colleagues assessed the donor cell chromosomal fragment loss on the genomic level by IRS (interspersed repeat sequence)-based quantitative PCR in macaque-hamster hybrids. The qPCR data displayed a significant loss of the donor cell fragment after ten passages (Karere et al. 2010). A mechanism proposed for the loss of donor cell chromosome fragments is that they cannot attach efficiently to the hybrid spindle apparatus. This model is supported by the centromere effect observed by Benham *et al* (Benham et al. 1989) and Goodfellow *et al* (Goodfellow et al. 1990), in which markers close to the centromeres of the donor cell chromosomes are often retained at higher than average frequencies. Centromeres are needed for proper chromosome segregation during meiosis. Nabhloz *et al* have suggested that chromosomes are lost by a slow and progressive loss in some instances over cell culture passages (Nabholz et al. 1969).

Since radiation hybrids are unstable, results obtained from the same clone can only be combined if DNA originating from the same cell culture passage is used. This requires that sufficient quantities of DNA should be obtained from one cell culture batch for mapping all markers and for sharing the RH panel resource between collaborating laboratories. In order to obtain sufficient DNA quantities, two strategies are widely used: large scale culture of the hybrid cells or another is whole genome amplification (WGA).

2.4.4 Whole genome amplification as an alternative approach to avoid large scale culture

During the first decade of RH mapping, large scale culture was the only method used to obtain the amount of DNA required for large-scale mapping. However, this was a time-

consuming and laborious process and a genome wide half-life of the donor DNA of 8.7 passages was reported in cell culture (Karere et al. 2010), meaning that the retention of donor markers decreased rapidly during cell culture. However, in the absence of any alternative method, most published panels, made in the late 1990s or the beginning of 2000s, have gone through large scale culture. Whole genome amplification (WGA) methods at the time were based on PCR, like for instance primer extension pre-amplification (PEP) (Telenius et al. 1992), or degenerate oligonucleotide primed PCR (DOP) (Zhang et al. 1992). Amplification products obtained by these PCR-based methods have a limited length, with a typical amplification fragment length of < 3 kb and an error rate of 3×10^{-5} .These methods also suffer from incomplete coverage and uneven amplification of genomic loci. Up to $10^{-2} \sim 10^{-4}$ and $10^{-3} \sim 10^{-6}$ fold amplification biases have been described using PEP and DOP-PCR methods, respectively (Silander and Saarela 2008). Therefore, these methods, although widely used for chromosome painting by FISH, were not adapted for producing DNA for RH mapping.

More recently, a new method of WGA was reported by Dean *et al*, using isothermal amplification by DNA polymerase phi29 called multiple displacement amplification (MDA) (Dean et al. 2002). The main difference between MDA and PEP or DOP is that the enzyme derived from the *Bacillus subtilis* bacteriophage phi29 allows the isothermal amplification at 30^{o}C and has strong strand displacement ability. Thus, the polymerase can produce DNA fragments as long as 70,000 nt on average and has DNA polymerase's associated 3'–5'exonuclease proofreading activity. Exponential amplification results through a hyperbranched DNA intermediate structure. Only 0.2% of genome loss has been reported by Lasken *et al.* (Lasken 2009). This method has been tested intensively for trait association studies (Pask et al. 2004), genetic disease research and for the sequence analysis of DNA (Berthier-Schaad et al. 2007) on homogenous DNA, and then Kerere et al have tested it on radiation hybrids (Karere et al. 2010). A number of experiments showed that MDA-amplified

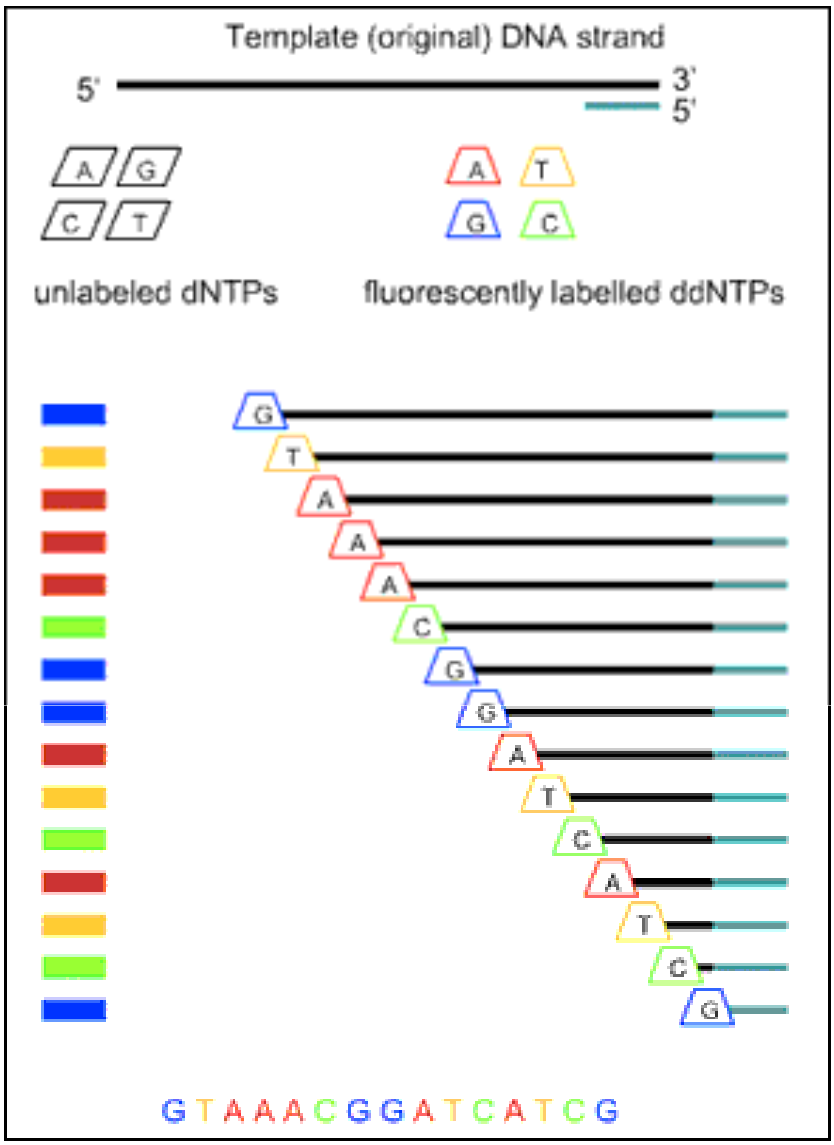


Figure I-11: principle of Sanger sequencing. Dideoxy nucleotides (ddNTP), each labeled with one of four fluorescent dyes are mixed with dNTP in the sequencing reaction. A random incorporation of a ddNTP stops the elongation of the DNA copy. The fragments obtained are separated by size by capillary electrophoresis the size and color of the fragments passing in front of a detector allows sequence determination.

genomic DNA is suitable for several common genetic analysis methods, including single nucleotide polymorphisms (SNP) genotyping, FISH chromosome painting, Southern blotting, restriction length polymorphism analysis, subcloning, and DNA sequencing (Lovmar and Syvanen 2006). Furthermore, a WGA European sea bass whole genome RH map has been published which indicates the feasibility of mapping whole genome amplified hybrids (Guyon et al. 2011).

2.5 Genome sequencing

2.5.1 The Sanger sequencing method

More than fifteen years elapsed between the discovery of the DNA double helix and the first experimental determination of a DNA sequence: the 24 bases of the cohesive ends of bacteriophage lambda (Wu and Taylor 1971). By the mid 1970s, sequencing methods were improved and two main ones were used. One was proposed by Maxam and Gilbert (Maxam and Gilbert 1977) and worked by chemical modification of DNA, followed by the cleavage at specific bases. However, this method used toxic chemicals and high amounts of radioactivity and was out of use after only a few years. The second was published by Sanger *et al* (Sanger and Coulson 1975) and has been prevalent for more than 30 years, with constant improvements. It was notably used for the first sequencing of an entire genome: the 5386 bp genome of the Phi X 174 bacteriophage (Sanger et al. 1977) and much later for the sequencing of the much larger human genome (Lander et al. 2001; Venter et al. 2001). The key to Sanger sequencing is the use of a polyacrylamide gel to separate by size the products of primed synthesis by DNA polymerase, with specific stops at the 4 possible bases A, C, G and T (Figure I-11). Although it required more steps than the Maxam and Gilbert technique, such

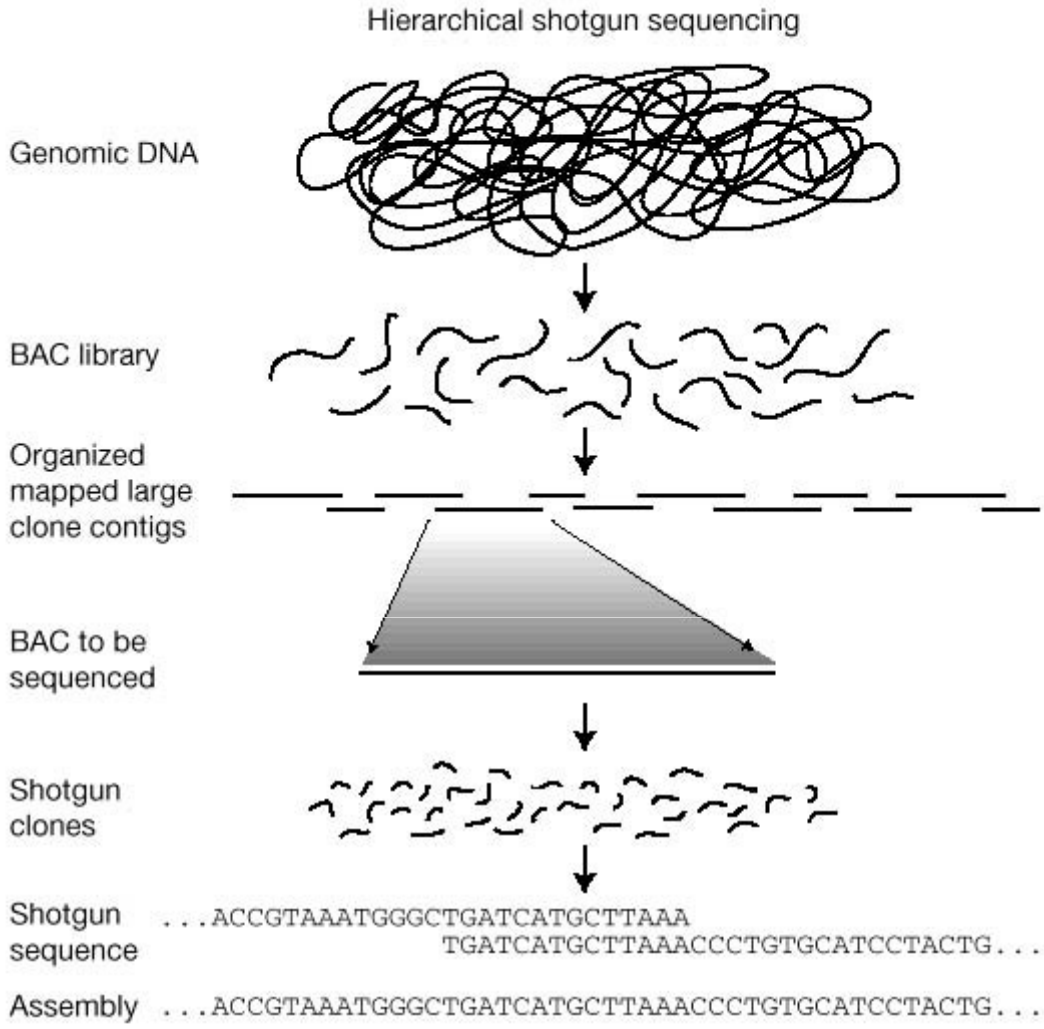


Figure I-12a: Hierarchical shotgun sequencing. Also called "clone-by-clone sequencing". This strategy follows a 'map first, sequence second' progression (E. Green, 2001). First a physical BAC contig map as built, from which individual mapped clones are selected for directed sequencing.

as the cloning of the fragments to be sequenced in vectors such as plasmids and therefore the culture of bacterial clones, it soon became the technique of choice (Hutchison 2007).

The main improvements of the Sanger sequencing method include improved chemistry allowing for longer reads, the replacement of the cloning step by PCR, and more importantly, the use of sequencing machines, performing the electrophoresis step in slab gels or more recently in capillaries, followed by the automatic reading of the fluorescently labeled DNA fragments.

Until 2005, Sanger sequencing (sometimes also referred to as capillary sequencing in reference to the latest generation of sequencing machines) was still the dominant technology used, with read lengths around 1,000bp and a per-base raw accuracy as high as 99.999% (Shendure and Ji 2008).

2.5.2 Strategies for whole genome sequencing of large genomes

There are two main strategies for whole genome sequencing and assembly (Green 2001): (1) The hierarchical shotgun sequencing approach also referred as “clone by clone” or “map-based” (Figure I-12a), follows a ‘map first, sequence second’ progression. The target DNA is firstly analyzed by a clone-based physical mapping methods, generally BAC contig mapping and after individual mapped clones spanning the region of interest are selected and sequenced. In this strategy, the process can be divided into a series of discrete and sequential steps: (i) map construction in which pieces of genomic DNA are cloned using a suitable host-vector system like BAC; (ii) clone selection in which selected clones representing a minimal tiling path across genome are chosen; (iii) subclone library construction, because a BAC clone insert are too large to be sequenced at once, the BAC clones are subcloned into vectors

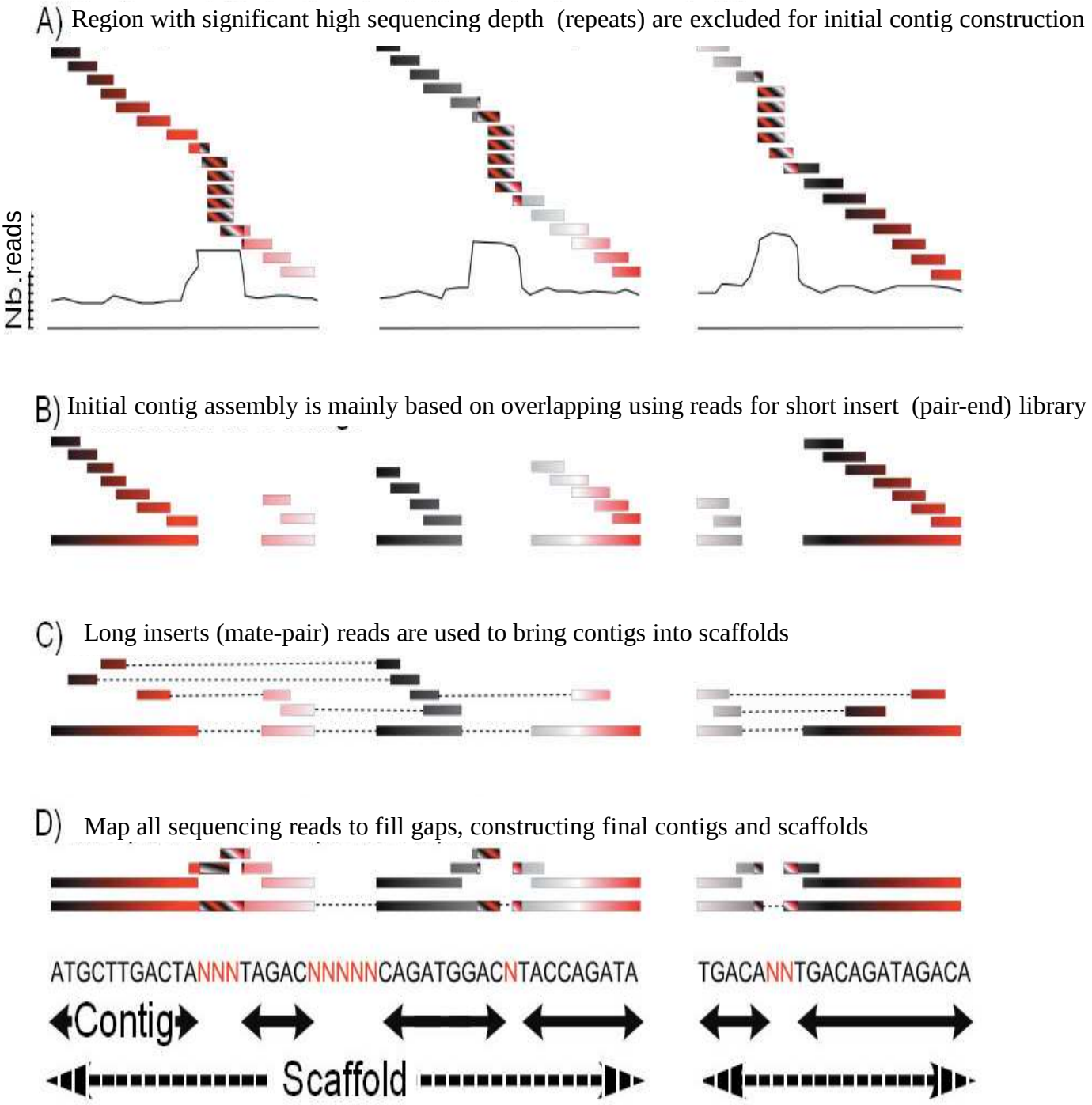


Figure I-12b: whole genome shotgun sequencing. Genomic DNA from one individual is sequenced directly. The assembly follows a stepwise strategy. First, initial contigs are assembled using short inserts reads; repeats are avoided by not extending into region with unusually high sequencing depth. Then, paired sequences: cosmid, fosmid or BAC ends in the case of Sanger sequencing or mate-pair reads in the case of NGS are used for building scaffolds. Finally all remaining reads are mapped back to the assembled scaffolds to fill the gaps. (Cited from A.Vignal **Etat actuel du séquençage et de la connaissance du génome des espèces animales**, INRA Prod. Anim. 2011, 24(4) 387-404)

containing smaller inserts such as plasmid or M13-based vectors, generating one library per BAC clone; (iv) randon shotgun sequencing and assembly of each BAC clone library. In this strategy, sequencing coverage at 3~5 fold is usually used for a wide range of analyses. Human (Lander et al. 2001) genome sequencing projects adopted hierarchy shotgun sequencing. (2) The whole genome shotgun sequencing approach which had been first applied mainly for the sequencing of repeat-poor and small genome such as viruses, bacteria and flies but also to larger genomes, with a higher repeat content, such as human (Venter et al. 2001). In this approach, the targeted DNA is fragmented into pieces of defined size and/or subsequently cloned into a suitable host-vector system. Sequencing reads are generated from both insert ends of a huge number of subclones so as to produce highly redundant sequence coverage across genome. The application of whole genome shotgun sequencing to eukaryotic genomes is more difficult than for bacteria or viruses, owing to their larger size and higher repeat content. Using pair-end reads and different insert sizes can help close gaps owing to repeat (Figure I-12b). And indeed, for repeat-rich genomes, different insert size libraries are crucial.

The two strategies mentioned above are not mutually exclusive but are often used in complement of one another. In a hybrid shotgun-sequencing strategy, sequencing reads are generated in both whole genome shotgun sequencing and hierarchical shotgun sequencing, capturing the advantages of both strategies. The whole genome shotgun provides a rapid insight into a genome and the hierarchy shotgun component simplifies the process of sequence assembly and minimizes the likelihood of serious misassemblies. This strategy can be used without a prior BAC contig map of the genome.

The chicken (Consortium 2004) genome sequencing project was an example adopting a hybrid strategy. The assembly was generated from a 6.6X coverage in whole genome shotgun reads, of a combination of plasmid, fosmid and BAC-end reads. Sequencing reads

were used to build first contigs which are sequences assembled without gaps and then supercontigs or scaffolds which are ordered and oriented contigs including estimates of gap sizes (principle in FigureI-12b). The draft assembly produced a N50 contig size of 36kb and a N50 supercontig or scaffold size larger than 7Mb, in which N50 statistic is defined as the largest length L such that 50% of all nucleotides are contained in contigs or scaffolds of size at least L. A BAC-based physical map was developed in parallel with the sequence assembly. Along with genetic map, combining BAC-based physical map and BAC-end reads helped grouping and orienting scaffolds, as well as assigning them to chromosomes. This larger-scale ordering and orienting of scaffolds by genetic and FISH maps, with assignment to chromosomes, creates the so called 'ultracontigs' or 'ultrascaffolds'.

However, no matter which strategy is adopted, the high cost of Sanger sequencing is a major problem for whole genome sequencing of species with large genomes such as most of higher eukaryotes. The Human genome project cost 3 billion dollars for the initial human genome sequencing (Lander et al. 2001) and the chicken genome, also sequenced by the Sanger method in 2004 cost almost 8 million dollars. The major cost in Sanger sequencing is not only due to sequencing reaction itself, but also to the library preparation, involving the growing of each sequence template in individual bacteria colonies, which is a time and labor-consuming step.

2.5.3 Next Generation Sequencing or parallel sequencing

In high throughput production pipelines, capillary sequencing machines such as the Applied Biosystem 3730, can read 96 sequencing reactions in a 2 hour run, producing 96 kb of sequence. Hence, whole genome sequencing project for large genomes require hundreds of

sequencing instruments and a large effort in library preparation. There are already several thousand prokaryote and eukaryote genomes sequenced including most model organisms and many agronomic species.

In many cases, to link phenotypes with genotypes, the current sequencing data available is not sufficient. Especially genome-wide association studies (GWAS), detection of genetic sweeps and epigenetics become more and more popular; species which have a reference sequence are resequenced for the detection of polymorphism and association with traits. Sequencing multiple species offers great help to comparative genomics which will provide insight into evolution and help detecting important sequences on the basis of conservation. Sequencing is a core technology in the development of genomics and genetics, and the high demand for low-cost sequencing drives the development of high throughput sequencing which parallelizes the sequencing process.

After four decades of gradual improvement, nowadays, the so-called Next Generation Sequencing (NGS) technologies replace Sanger sequencing for projects needing large amounts of data. Several milestones should be noted during NGS evolution: (1) Ronaghi *et al* (Ronaghi et al. 1996) announced the method of pyrosequencing in 1996, which is used by 454 sequencing; (2) Lynx Therapeutics, merged with Solexa Ltd. which is now merged with Illumina have published and marketed the method of 'Massively Parallel Signature Sequencing' (MPSS) in 2000 (Brenner et al. 2000); (3) 454 Life Science commercialized a parallel version of pyrosequencing in 2004; (4) Shendure *et al* developed a method named Polony sequencing (polymerase colony) in 2005 and ultimately incorporated it into the Applied Biosystem SOLiD platform (Shendure et al. 2005); (5) Bentley *et al* announced and published the sequencing of a flow-sorted human X chromosome by the Illumina Solexa

sequencing technology, proving the feasibility of using ultra short read for assembly and structure variant detection (Bentley et al. 2008).

The main innovation of NGS, as compared to Sanger sequencing, is the parallelisation of the process, allowing between a few thousands and up to millions of sequencing reactions to be processed simultaneously. The main aspects are (1) complete libraries are used directly for sequencing and DNA fragments are amplified in parallel, thus tremendously reducing the cost and complexity when compared to Sanger sequencing in which individual *E.coli* clones were picked and grown; (2) templates from the libraries are processed in parallel on an immobilized surfaces where they are enzymatically manipulated by a single reagent volume; (3) sequencing is done in cycles in which only one base of each template is interrogated, the number of cycles determining the read length. The three main NGS technologies available on the market use different approaches for library construction, template immobilisation and sequencing reaction, but the basic principles remain the same. NGS also have some drawbacks compared with Sanger sequencing: (1) sequence reads produced by NGS (100 bp for Illumina, 500 bp for 454) are shorter than Sanger sequencing reads (1,000 bp) and have a higher error rate, making the sequence assembly more problematic; (2) the pairing of reads in Sanger sequencing is limited by the size of the DNA fragments that can be inserted in cloning vectors, ranging from 1-2 kb or less up to 100-200 kb (plasmids, fosmids, BAC), whereas pairing of reads is limited to 10 kb with NGS, limiting the average assembled scaffolds length and leading to more difficulty in segmental duplication and copy number variation detection; (3) as a consequence, higher sequencing depth is required for assembly. Nevertheless, the sequencing and de novo assembly of a Chinese individual (Li et al. 2010b) proved the feasibility of sequencing and assembling whole genomes by NGS despite short sequencing reads. Many species were sequenced and/or resequenced by NGS since then, such as the giant panda (Li et al.), the silk worm (Xia et al. 2009), the cucumber (Huang et al. 2009), the

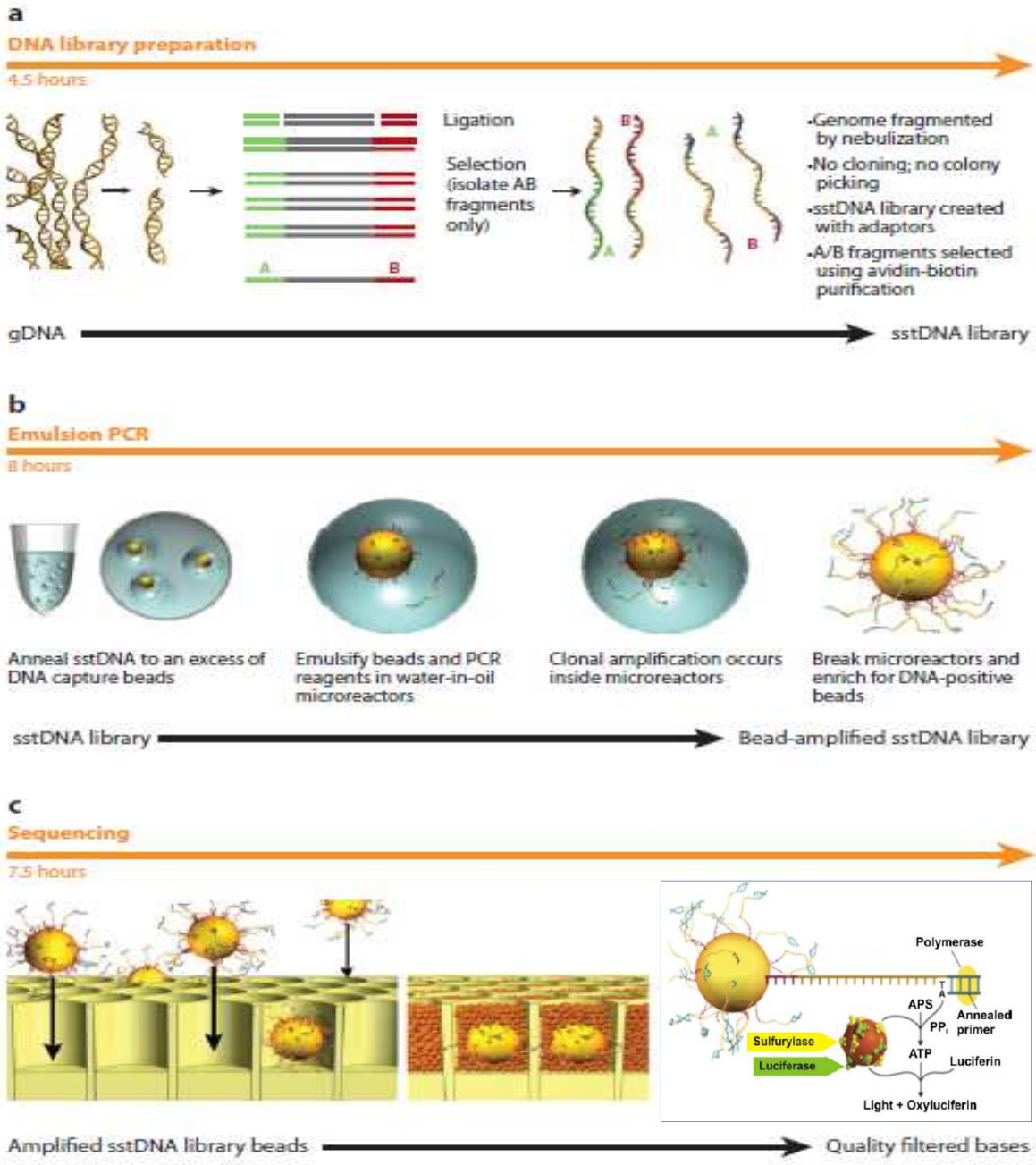


Figure I-13: Roche 454 sequencing workflow. a: library prepartion, DNA is randomly sheared and adaptors ligated to the ends. b: DNA template is amplified by emulsion PCR. DNA and beads are diluted and mixed with a ratio such as a maximum of one DNA molecule is attached to a bead. An emulsion is created so that individual beads are trapped into microreaction chambers. PCR allows then the clonal amplification of DNA molecules attached to beads. c: Beads bearing amplified DNA are loaded on the PTP plate into microwells in which the sequencing reaction takes place. One of four dNTPs is added sequencially in the reaction. Each incorporation event results in releasing a pyrophosphate which is thereby converted into an ATP with the presence of APS and sulfurylase. The ATP provides energy to luciferin which thereafter generates visible light under the catalysis of luciferase. Thus when one of the four possible nucleotides is added in the reaction, light is emitted only from the microwells in which it is incorporated. A picture of the PTP plate is taken at each cycle, capturing the information in parallel from all reactions simultaneously. (http://www.454.com)

chicken (Rubin et al. 2010; Ye et al. 2011) and the duck (Huang *et al.*, in prep). An analysis performed by Warren *et al* (Ye et al. 2011) showed that NGS can somehow get more sequences from chicken than Sanger sequencing does. To date, three NGS technologies which are Roche 454, Illumina Solexa and ABI SOLiD dominate the sequencing market.

a. Roche 454

This next-generation sequencer was first to achieve commercial introduction in 2004 and uses an alternative sequencing technology known as pyrosequencing (Ansorge 2009; Mardis 2008). The first instrument named 454 GenomeSequencer (GS) FLX was introduced in 2005. The workflow is summarized in Figure I-13 and contains mainly three steps: (1) DNA library preparation; (2) Emulsion PCR to amplify DNA template and (3) sequencing.

DNA/RNA is fragmented by nebulization and is subsequently ligated with adapters allowing the binding to oligonucleotides on the surface of beads. Ligated libraries and beads are mixed in proportions such as only one DNA fragment per bead is fixed. An emulsion is then created with oil in which each single bead is incorporated into one droplet, which behaves as a micro reactor. Emulsion PCR is then carried out to amplify the DNA templates and each bead carries now copies of the single DNA molecule that was bound, allowing for sufficient light signal intensity for reliable detection in the sequencing process and also for preventing template cross contamination. When PCR cycles are complete, beads are treated with denaturants to remove the untethered strands and then subjected to a hybridization-based enrichment for amplicon-bearing beads. A sequencing primer is hybridized to the universal adapter (Ansorge 2009). Then each bead with its amplified fragment, *Bacillus stearohermophilus (Bst)* polymerase and single-stranded binding protein placed in the PicoTiterPlate (PTP) plate which contains millions of etched picoliter wells, created from

glass fiber bundles. During the sequencing process, one of the four possible dNTP is introduced in each cycle. A pyrophosphate is released if there is one nucleotide incorporated, and then released pyrophosphate incorporates with an Adenosine 5'-phosphosulfate (APS) into an Adenosine triphosphate (ATP). Via ATP sulfurylase and luciferase, incorporation events immediately drive the generation of a burst of light which is detected by CCD which is bonded to the other side of PTP plate.

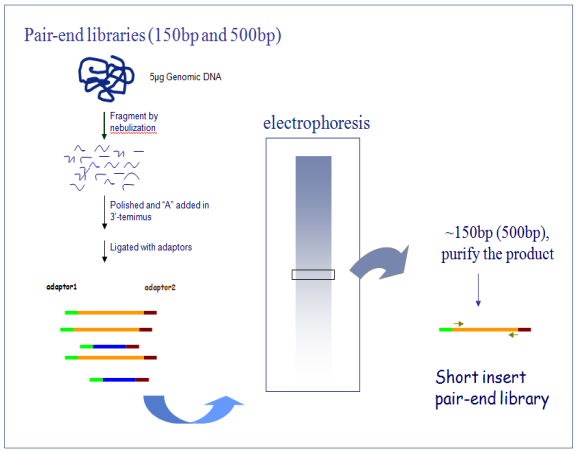
There is a major drawback in 454 sequencing concerning homopolymers like AAAAA or CCCC, due to the reason that the length of homopolymer is determined by the signal intensity. Therefore the dominant error type from 454 platform is insertion-deletion related to homopolymers. But compared with other platforms, the key advantage is read length. For the moment, the GS FLX Titanium XL+ system can read up to 1000bp.

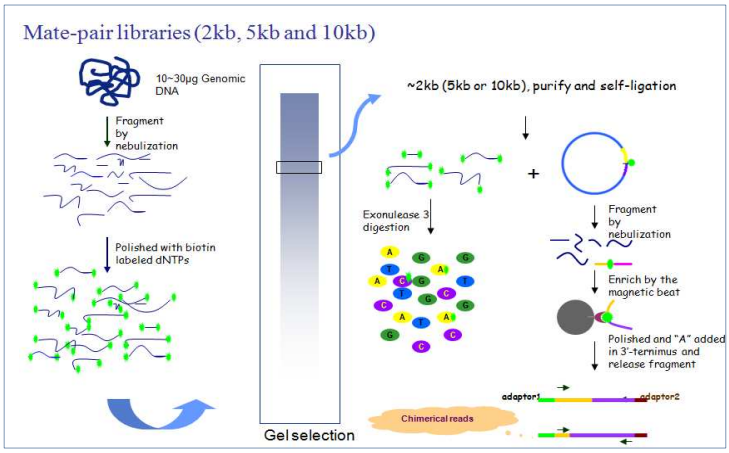
b. Illumina Solexa

The Illumina Solexa sequencing platform was commercialized in 2006. This platform has its origins in work by Turcatti and coworkers (Fedurco et al. 2006; Turcatti et al. 2008). They solved the most two tough obstacles for Solexa platform. One breakthrough is that they used Benzene-1,3,5-triacetic acid (BTA) to attach 5'-aminated DNA primers and templates on an aminosilanized glass surface for subsequent generation of DNA colonies by *in situ* solid-phase amplification. By this innovation, the primers on the surface of the glass flow cell are more stable and heat-resistant. The other breakthrough is that they used a 3'-OH unprotected cleavable fluorescent 2'-deoxynucleotides to block the next incorporation events in DNA synthesis.

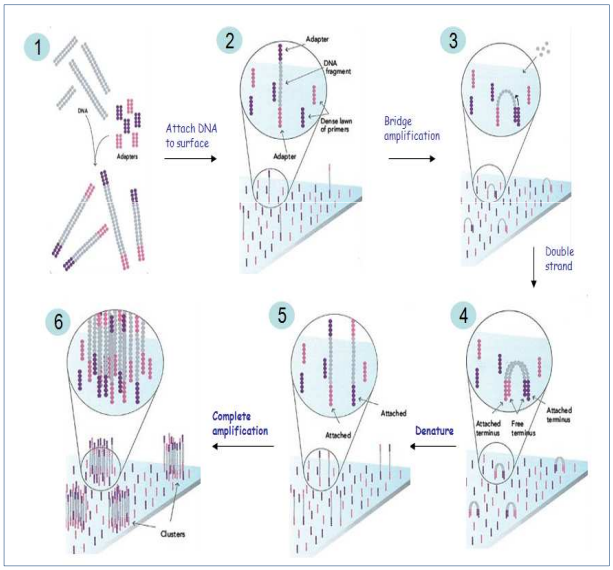
Illumina

Library preparation





Cluster generation



Sequencing

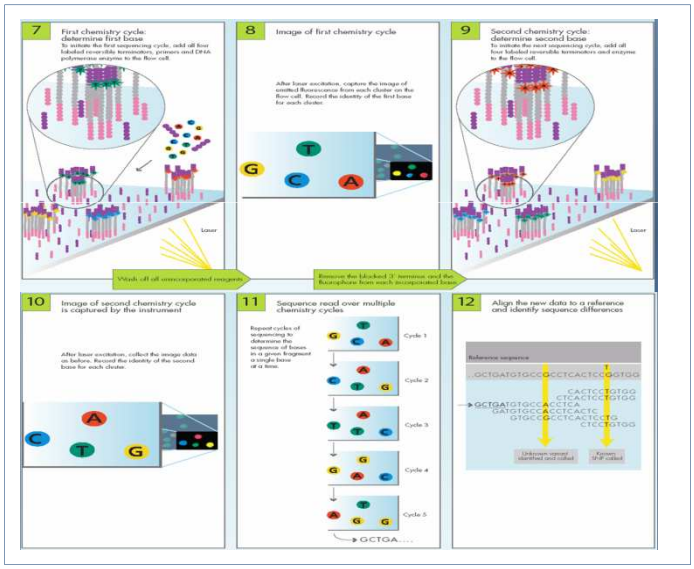


Figure I-14: overview of Illumina sequencing workflow. Top: library preparation involves DNA fragmentation and adaptor ligation to construct pair-end libraries. The construction of mate-pair libraries is more complex, involving several ligation and purification steps. Sequence pairs further apart can be obtained, but with a small rate of false-pairing. Pair-end libraries are more reliable than mate-pair libraries. Single molecule templates are immobilized onto the surface of a flow cell and cluster generation is done by bridge PCR, in order to amplify the signal intensity. Sequencing by synthesis reads one base at a time by using 3'-blocked reversible terminator. The 4 bases are labeled with different fluorochromes. A picture of the flowcell is taken at each cycle, indicating which base is incorporated at each position on the flowcell. (http://www.illumina.com)

The principle of Solexa sequencing is sequencing by synthesis with novel reversible terminator nucleotide for the four bases, each labeled by a different fluorescent dye. In the Solexa sequencing workflow (Figure I-14), there are three steps: (1) library preparation which is similar to 454 platform; (2) cluster generation which uses bridge PCR to amplify locally the single DNA strands attached to the glass plate (flow cell) and (3) sequencing by synthesis.

In the library preparation process, there are two categories of libraries: paired-end libraries and mate-pair libraries. In practice, paired-end libraries refer the short insert size library (150 ~ 800bp), whereas mate-pair libraries concern longer insert sizes (2kb ~20kb); both ends of DNA in both library types can be sequenced. Briefly, for short-insert library generation, genomic DNA is fragmented by nebulization with compressed nitrogen gas. Then the DNA fragments are polished at the both ends and an "A" base is added to the ends. The DNA adaptors with a single "T" base overhang at 3'-end are ligated to the above products. Then the ligation products are purified on an agarose gel, the required size band are excised and purified for sequencing. For the mate-pair libraries, genomic DNA is fragmented by nebulization, the DNA fragments are polished by biotin labeled dNTPs and the required sizes are selected on agarose gel. The purified DNA fragments are circularized by self-ligation, so the two ends of the DNA fragment are merged together; any remaining linear DNA fragments are digested by a DNA Exonulease. The circularized DNA are fragmented again by nebulization, followed by enrichment of the "merged end" with magnetic beads and biotin/streptavidin, then the ends are polished and "A" base and adaptors added for sequencing.

Cluster generation is performed to enrich the templates. DNA molecules from the library are denatured and attached to the surface of a flow cell (which contains 8 lanes) where there are dense lawns of primers whose sequences are complementary to the adaptors. After the single-strand DNA molecules binding to the primers in the flow cells, unlabeled

nucleotides and DNA polymerase are added to initiate solid-phase bridge amplification. The enzyme incorporates nucleotides to build double-strand bridges on the solid-phase substrate. Then the double-strand DNA is denatured and subsequent cycles are performed to perform an *in situ* PCR. Complete amplification allows the generation of dense template clusters.

The last step of the workflow is the sequencing. A flow cell containing millions of unique clusters is loaded into the sequencer for automated cycles of extension and imaging. The first cycle of sequencing consists in the incorporation of a single fluorescent nucleotide, followed by high resolution imaging of the entire flow cell. These images represent the data collected for the first base, each of the 4 possible bases having its specific fluorescence wavelength. Any signal above the background identifies the physical location of a cluster and the fluorescent emission identifies which of the four bases is incorporated at that position. To initiate the first sequence circle, all four labeled reversible terminator, sequencing primer and DNA polymerase are added into the flow cell. All unincorporated reagents are washed away and the image of emitted fluorescence from each cluster is captured after laser excitation. The blocked 3' terminus and the fluorophore from each incorporated base are then removed. To initiate the second sequence cycle, all four labeled reversible terminators and DNA polymerase are added again and the process above is repeated until the end of the run. Base calls are derived with an algorithm that identified the color emission over time for each cluster position on the flow cell.

The HiSeq2000 system can produce 300Gb of sequence per run with read lengths up to 150bp. In the Illumina Solexa platform, the main errors are substitutions rather than insertion-deletions. Average raw error rates are on the order of 1-1.5%, but higher accuracy bases with error rates of 0.1% or less can be identified through quality metrics associated with each base call (Shendure and Ji 2008).

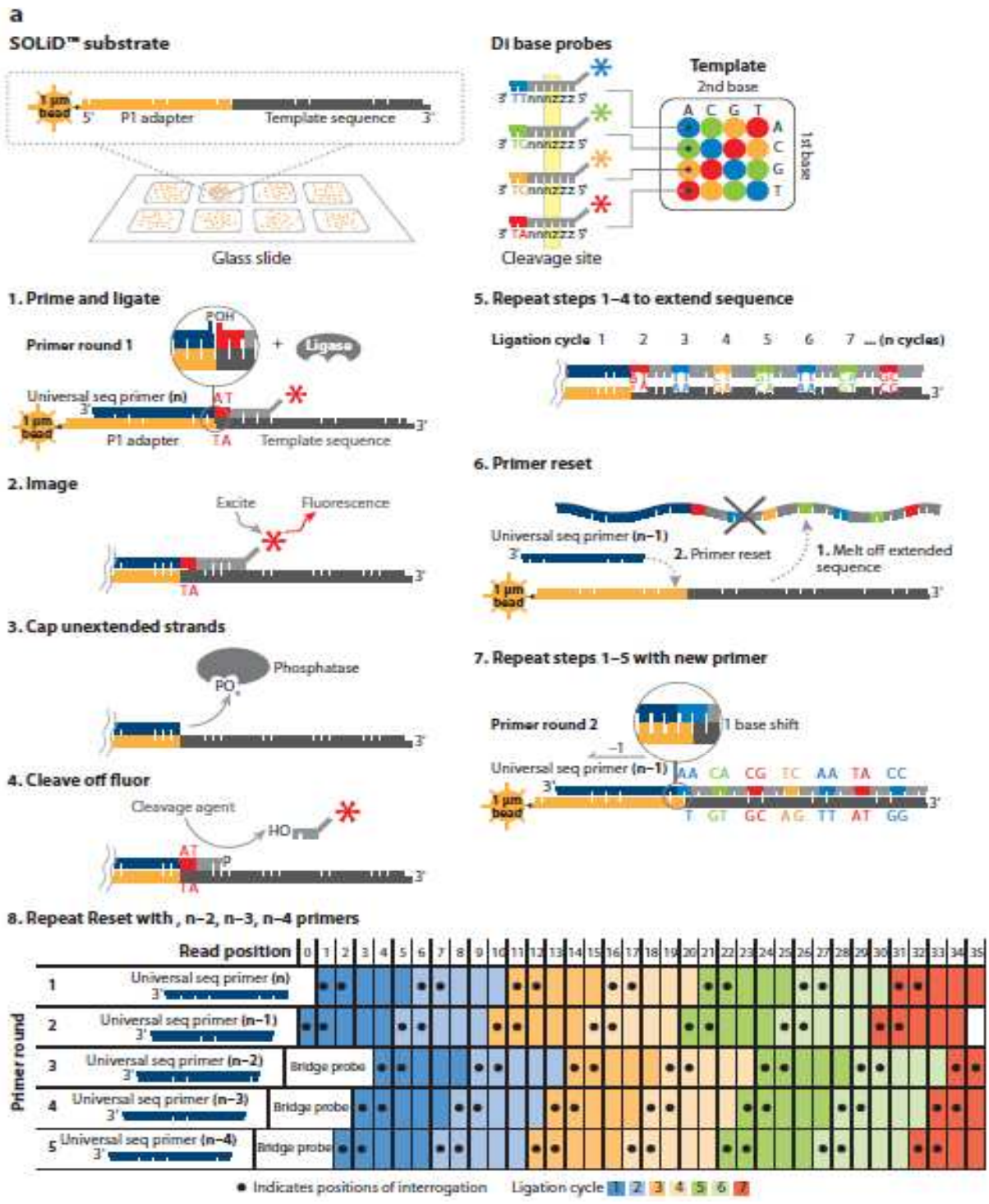


Figure I-15: Solid sequencing principle. Similarly to the Roche workflow (Figure 13), DNA is fragmented and ligated to adaptor P1, emulsion PCR is performed to amplify template and the beads are deposited on the glass slide. The sequencing reaction works by the ligation of octamers, instead of the usual incorporation of dNTP by a polymerase. The two first bases of the octamers are used to interrogate the sequence to be read. Four colours are use for 16 di-nucleotides and the sequence is deduced from the knowledge of the first base. After each incorporation, cleavage takes place at the 5th base. Then repeats the incorportaion until the reaction is finished, then product is removed and the template is reset with a primer complementary to the n-1 position for a second round of ligation cycles. (http://www.appliedbiosystems.com)

c. ABI SOLiD

The ABI SOLiD was first introduced in 2007 and its principle is sequencing by ligation. Like for the other two platforms, three main steps can be characterized: library preparation, cluster generation and sequencing. DNA library preparation is also an *in vitro* process in which, similarly to the Roche 454, DNA is fragmented, ligated to designed adaptor, attached to magnetic beads, emulsion PCR is applied to generate the clusters and beads are plated onto a solid support with microwells. Unlike the other two platforms, SOLiD uses DNA ligase rather than a polymerase and a unique approach to sequence the amplified fragments. The details are shown in Figure I-15. A universal primer (n) complementary to the adaptor sequence is hybridized to the array of amplicon-bearing beads. A set of structured fluorescently labeled octamer mixtures are involved to decode the sequence. In these octamers, the two first bases are used to characterize di-nucleotides and are each characterized by one of four fluorescent labels at the end of the octamer. After ligation and signal detection, the ligated octamers are cleaved at 5^{th} base to release the fluorescent labels, and then hybridization and ligation cycles are repeated, leading to the identification of di-nucleotide unit with intervals of three nucleotides (Figure I-15). After the DNA synthesis with universal primer (n), the newly synthesized strand is denatured and washed off. Then another universal primer (n-1) which is one base less than universal primer (n) and fluorescently labeled octamer mixtures are added into reaction again to repeat the procedure above. Altogether there are five primer rounds which contain 5 different universal primers from n to n-4 which means each base has been decoded twice, therefore improving the accuracy of the reads.

The current read length achieved is 75bp, for the 5500xl system and each instrument can produce approximately 15Gb per day. Each base is queried twice due to di-base decoding so that the error rate is reduced.

2.5.4 Comparison and Conclusion

In term of costs among these three sequencing technologies discussed above, Illumina Solexa and ABI SOLiD have a lower cost per base sequenced compared with Roche 454, but the read length and average acurracy (Harismendy et al. 2009; Margulies et al. 2005) in Roche 454 can now almost rival with Sanger sequencing. Considering the error type in these three NGS platforms, the main error type for the Roche 454 platform is insertion-deletions, due to its weakness in the determination of the number of bases in homopolymers, and for the other two platforms, the main error type is substitutions (Dohm et al. 2008).

Roche 454 and ABI SOLiD both use emulsion PCR to generation clusters, which can be cumbersome and technically challenging. In SOLiD platform, it is possible that sequencing on high density array of very small beads may represent the most straightforward opportunity to achieve extremely high data density, simply because 1μm beads physically exclude one another at a spacing that is on the order of the diffraction limit. Furthermore, high resolution ordering of 1μm bead arrays may enable the limit of one pixel per sequencing feature to be closely approached (Shendure and Ji 2008). The output of the reads from ABI SOLiD differs from Illumina solexa and Roche 454 as well because of the sequencing by ligation principle.

According to a case study by Ye *et al* (Ye et al. 2011), the same red jungle fowl individual that had been sequenced sequenced in 2004 by Sanger sequencing was resequenced by Roche 454 and Illumina Solexa platforms. A notable problem in the chicken sequencing is that the smallest 10 microchromosomes were missing in the final assembly. More than 31Mb of new sequence data was obtained by NGS. Comparing the novel sequences obtained from these two platforms, showed that those obtained from Illumina Solexa platform had higher GC content and Roche 454 sequences contained more contaminated sequences before

contamination removal. Partial novel sequences have BLAST hits with novel sequenced obtained in recently sequenced BAC clones selected for the finishing process, therefore NGS platforms may have better performance in whole genome sequencing for covering regions missed by Sanger sequencing. Suzuki *et al* (Suzuki et al. 2011) resequenced the same straine of *E.coli* DH1 recently, and the result showed that Illumina Solexa and ABI SOLiD have a relatively higher proportion of unusable reads among these three NGS technologies, but the low quality reads can be trimmed without a doubt.

These NGS technologies have their own merits and drawbacks, and each technology has been successfully applied to whole genome sequencing, resequencing, DNA methylation analysis (Li et al. 2011) and transcriptomics (Mortazavi et al. 2008). Along with the 1000 Genomes project, developments of metagenomics and technological optimization, NGS is now prominent. Given the state of flux, the near future will be an era of NGS and its extents.

2.5.5 Consequences of the NGS on genome assembly strategies

As described above, the sequencing reads in NGS are shorter than with Sanger sequencing and the library insert size that can be sequenced are also smaller in NGS. The current largest insert size for NGS is around 20kb (even though some company announced that maximum inserts could be up to 40kb) whereas BAC libraries can have insertion up to 180kb in length and BAC clones can be sequenced from both ends by Sanger sequencing. In sequence assembly, the principle is similar in NGS and Sanger sequencing, although there are some differences in the algorithms adopted. Contigs are build based on overlap information from the pair-end library and not extended to the regions which have too low, to avoid sequencing errors, or too high, to avoid repeats, sequencing depths. Using a stepwise strategy, mate-pair (large insert library) reads assemble the contigs into supercontigs or scaffolds

(Figure I-12b). Finally all reads are used to close the gaps when possible. In the scaffolding process, the size of the mate-pair libraries influences the length of N50 or average scaffold or supercontig length. The larger the insert size, the longer the scaffolds. Current whole genome sequencing projects by NGS now concern species for which there is a lack of detailed or satisfactory supplementary long range mapping data. This makes the construction of ultracontig or ultrascaffold and especially the chromosomal assignment much more difficult.

Due to the smaller insert size of libraries, the scaffolds are smaller and hence the number of scaffolds is larger. For instance, for panda (Li et al. 2010a), which is the first mammal sequenced by NGS, the assembly gives N50 contig and N50 scaffold sizes of 39kb and 1.28Mb respectively.

2.5.6 Third generation sequencing

NGS methods have changed whole genome sequencing projects into routine procedures and have been adapted to other areas, such as transcriptome sequencing and epigenetics. However, in second generation sequencing platforms, as described above, templates are amplified by PCR before the sequencing step to make the light signal strong enough to be detected. The use of PCR is problematic for two reasons. First, amplification efficiency varies according to the property of template, for example the GC content, and thus introduces biases; second, errors are introduced in the process of PCR amplification and in a recent human genome resequencing for breast and colorectal cancer approach, it was found that PCR errors alone account for about one third of initially detected mutations (Sjoblom et al. 2006). The fidelity of PCR polymerases is reported to vary between $0.5x10^{-4}$ and $1.0x10^{-4}$, which is a substantial error rate for amplifying a single template (Barnes 1992). To overcome this, the ultimate miniaturization into the nanoscale and the minimal use of the biochemicals,

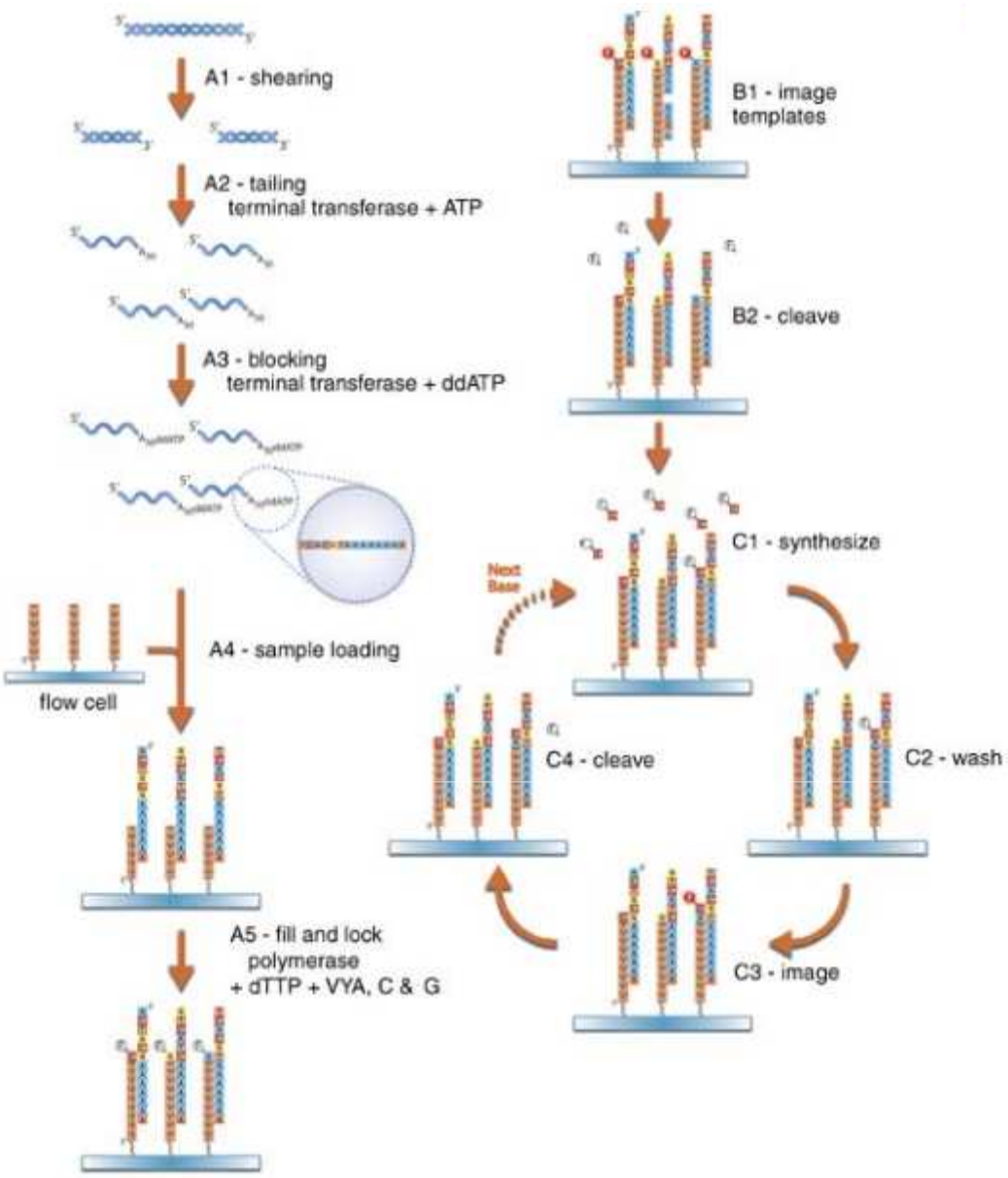


Figure I-16: Helicos Heliscope sequencing. A1-A3, template preparation. A poly(A) tail is added to the 3'-end of the template and then a ddATP is added to block the 3' terminus. A4-A5: template loading and unpaired dA are filled by dTTP and the reaction is stopped by virtual terminator nucleotides (ATP, CTP, GTP). B1-C4: sequencing by synthesis in a "wash and scan" manner, one base at each cycle. Pictures at each step are taken by CCD cameras via a confocal microscope. Thus, the principle is similar to the Illumina sequencing, but without the bridge PCR for signal amplification.(Ozsolak et al. 2009. **Direct RNA sequencing**. *Nature* **461:** 814-818)

would be achievable if the sequence could be determined directly from a single DNA molecule, without the need for PCR amplification and its potential for distortion of abundance levels. This sequencing platform sequence a single DNA molecule is now called as the third generation sequencing (TGS) technology (Pareek et al. 2011; Schadt et al. 2010).

a. Helicos Bioscience single molecule sequencing

One of the first techniques for sequencing from a single DNA molecule was introduced by Braslavsky et al (Braslavsky et al. 2003) and licensed by Helicos Biosciences at the first commercial single-molecule DNA sequencing in 2007. The Heliscope sequencer uses true single molecule sequencing (tSMS) technologies which was first applied for sequencing the M13 virus genome in 2008 (Harris et al. 2008) and has been successfully applied on direct RNA sequencing in 2009 (Ozsolak et al. 2009). The principle is described in Figure I-16. The library preparation involves DNA shearing and the addition of a poly-A tail to the fragmented DNA using *E. coli* poly(A) polymerase I (PAPI). The elongation step is blocked by introducing 3' deoxyATP to the polyadenylation reaction shortly after the start of the tailing reaction.This poly(A) tail is used for attaching the DNA fragment on the sequencing support, which is composed of poly-T oligonucleotides covalently anchored onto the surface of a flow cell at random positions. These oligomers are first used to capture the template DNA and then serve as primers for the sequencing step. This sequencing technique relies on stepwise synthesis in cycles in which one of the four nucleotides is added. The sequencing by synthesis reaction is performed using a modified polymerase and proprietary fluorescent nucleotide analogues, called Virtual Terminator nucleotide (VT), containing a fluorescent dye and which are chemically cleavable, allowing stepwise sequencing (Ozsolak et al. 2009). Observation of

single the molecule signals is accomplished by CCD cameras via confocal microsopy. The workflow is diagrammed in Figure I-16.

.

It is announced that the capacity of the Heliscope Sequencer is approximately 28 Gb in a single sequencing run of about 8 days. It can generate short reads with a maximal length of 55bp (Pareek et al. 2011). The sequencing error rate is reported around 4% and all the errors are single base errors because the sequencing is in "one base at a time" manner. The dominant error is deletion (2-3%) due to failures in detection of base incorporations; while the insertion rate is 1-2% probably caused by failing in rinsing VT analogues from the flow cell between each addition cycle. The substitution rate is 0.1-0.2%.

In summary, the principle is similar to the Illumina NGS sequencing, but without the bridge-PCR step.

b. Ion Torrent's Semiconductor Sequencing

Ion Torrent System Inc. has announced the semiconductor sequencing technology in 2011. The difference from other sequencing technologies is that the base calling is not based on optical methods and the detection of fluorophores. The principle of semiconductor sequencing technology used is that the base calling process is determined by the detection of a voltage change due to the fact that a hydrogen ion (or proton) is released after each nucleotide incorporating into the nascent DNA strand, resulting in a change of pH. The sequencer has no optical component and is comprised primarily of an electronic reader board interfaced with the chip, a microprocessor for a single processing, and a fluidics system to control the flow of reagent over the chip (Rothberg et al. 2011).

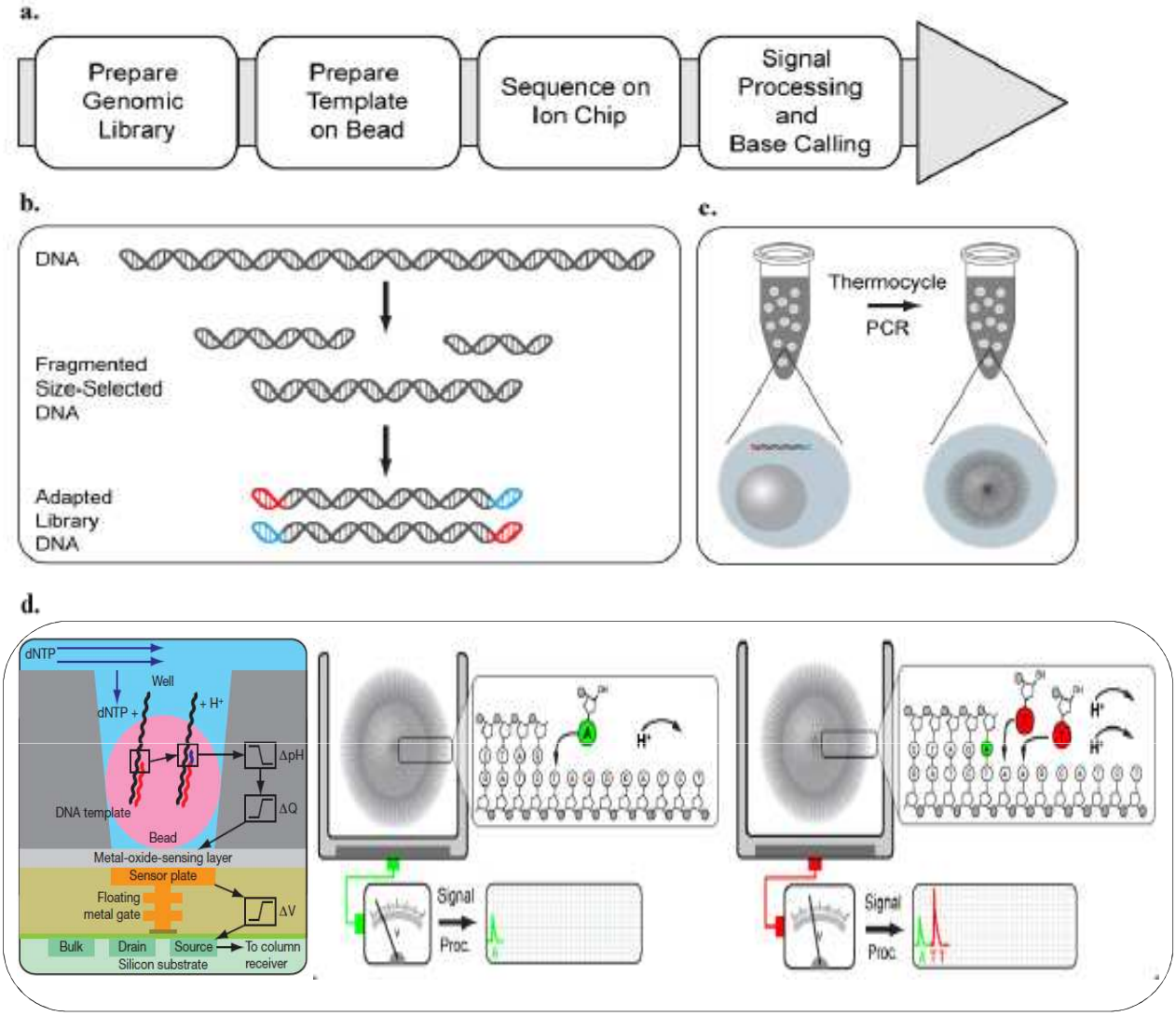


Figure I-17: Ion Torrent semiconductor sequencing. a: overview of the workflow; b: DNA library preparation. DNA is fragmented in a defined size and then ligated with adaptors; c: DNA template is amplified on beads by emulsion PCR; d: beads are deposited on a flow cell and DNA is sequenced by sequencial addition of dNTP, one of the four possible bases at a time. Each incorporation event causes the release of a proton, thereby influencing the pH of the neighboring solution. This in turn changes the surface potential of the metal-oxide-sensing layer leading to a change of voltage. The change of voltage is positively related to the number of released proton and thus the number of nucleotides incorporated. (Rothberg et al, 2011, **An integrated semiconductor device enabling non-optical genome sequencing**. *Nature* **475:** 348-352.

This technology is not true single molecule sequencing because it requires emulsion PCR to amplify the templates to amplify the signal. The library preparation can be briefly summarized as following: genomic DNA is fragmented and ligated with adapters and then adaptor-ligated libraries are clonally amplified onto beads like for the 454 pyrosequencing library preparation; and then emulsion PCR is performed to amplify the templates to achieve a high signal-to-noise ration.

In the sequencing step, similarly to 454 pyrosequencing, all four nucleotides are provided in a stepwise fashion during an automatic run and the sequencing run is a "cleavage and washing" manner. When the nucleotide is incorporated results in the net liberation of a single proton (or hydrogen ion) during that flow which thereby produces a shift in the pH of surrounding buffer proportional to the number of nucleotide incorporated. Then the pH change will be detected by the sensor on the bottom of each well, converted to a voltage and digitized by off-chip electronics (Figure I-17). This process eliminates the need for light, scanning and cameras to monitor the sequencing by synthesis process, thereby simplifying the overall sequencing process, dramatically accelerating the time to result, reducing the overall size of the instrument, and lowering cost to make DNA sequencing more generally accessibly.

Rothberg *et al* have characterized the performance of this technology by sequencing three different bacterial genomes. It is reported that the per-base accuracy was observed to be 99.569% within the first 50 bases and 98.897% within the first 100 bases; and the homopolymer of lenth 5 is 97.328% and higher than that of pyrosequencing-based method. This technology has allowed the routine acquisition of 100-based read lengths and perfect read length exceeding 200 bases. In their approach, 20-40% of the sensors in a given run yield mappable reads and the failure of the other sensors is probably due to incomplete loading of

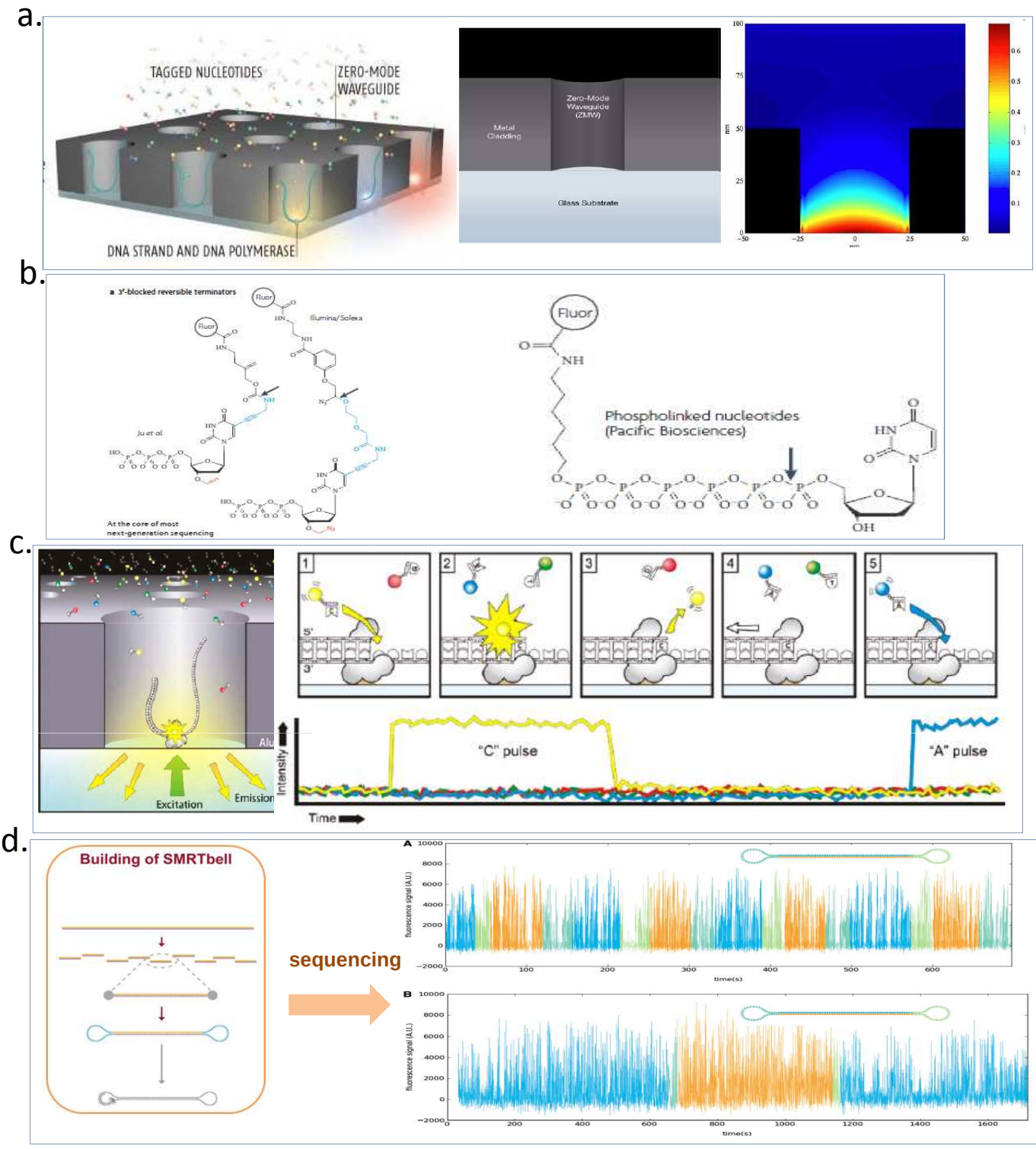


Figure I-18: Single Molecule Real Time (SMRT) sequencing., a: the structure of zero-mode waveguide(ZMW) in which the reaction takes place. Left: a schema of SMRT cell, with the single DNA molecules (blue) being processed; middle: structure of a single ZMW which is a small hole made by etching glass surface; right: the heatmap of the laser light passing through ZMW showing the light decayed dramatically;[a1] b: the modified nucleotide used in Illumina sequencing and SMRT sequencing. Left: the nucleotide used in Illumina platform in which the fluorophore is linked to the base; right: the nucleotide structure in SMRT sequencing, in which the fluorophore is linked to the phosphate chain; c: the schema of SMRT sequencing. In a single ZMW, DNA polymerase is immobilized on the bottom and laser light comes from beneath. The free nucleotides diffusing in the ZMW only stay for a few microseconds, while the nucleotides being incorporated are captured by the polymerase for several milliseconds, resulting into a wider pulse. The succession of wide pulses determines the sequence of nucleotide incorporation and therefore the DNA sequence. (http://www.pacificbiosciences.com)

the chip, poor amplification of the template on the beads or to beads bearing multiple templates (Rothberg et al. 2011).

c. Single Molecule Real Time Sequencing (Pacific Bioscience)

Single Molecule Real Time (SMRT) sequencing was licensed by Pacific Bioscience Inc. in 2009 and is reported as true single molecule sequencing in real time. The principle of this technology relies on single molecule real time sequencing by synthesis on a zero-mode waveguide (ZMW)-containing SMRT cell (Figure I-18). Differently from the ion semiconductor sequencing and Helicos true single molecule sequencing, SMRT sequencing achieves sequencing in real time and allows long sequencing reads which can be up to 10,000 bases long (Eid et al. 2009). SMRT sequencing technology has the advantages of shortening the time for obtaining results, of avoiding PCR amplification of the template and allows for long read length. Those advantages are achieved by two principle components: ZMW and fluorescence-labeled phospholinked nucleotides (Korlach et al. 2010).

The ZMW nanostructures consist of dense arrays of holes which are approximately 100nm in diameter, fabricated in a 100nm metal film deposited on a transparent substrate (Foquet et al. 2008; Levene et al. 2003). Each ZMW becomes a nanophotonic visualizable reaction chamber for observing a single nucleotide incorporation event, providing a reaction volume of ~100 zeptoliters (10^{-21} L). As the diameter of the ZMW is of three orders of magnitude smaller than the wavelength of fluorescence, the intensity of fluorescence from the free nucleotides in the reagent decreases dramatically when observed from the bottom of the reaction chamber by diffraction-limited confocal microscopy. The small size of ZMW prevents visible laser light which comes beneath the transparent substrate and has a

wavelength of 600nm from passing entirely through the ZMW. Rather than passing through, the light exponentially decays as it enters in the ZMW, and only the bottom 30nm of the ZMW becomes illuminated. In addition, the DNA polymerase is immobilized at the surface of the ZMW by streptavidin and biotin interaction. Therefore, it is possible to observe single nucleotide incorporations undergoing at the bottom of the reaction chamber or ZMW. Thereafter the fluorescent signal from each single chamber is transmitted and collected by the optical systems beneath the ZMW.

In addition to reducing the number of labeled nucleotides present inside the observation volume, the highly confined volume results in drastically shorter diffusional visitation times. This enables better temporal differentiation between events involving diffusion of labeled nucleotides through the ZMW which typically lasts for a few microseconds and incorporation events which lasts for several milliseconds, therefore, the diffusion events can be easily distinguished (Korlach et al. 2010).

ZMW only resolves the difficulties of observing single molecules during sequencing. The higher speed in sequencing reaction is achieved by the use of dye-labeled terminal phosphate-linked nucleotides. Several of the sequencing by synthesis schemes utilize nucleotides with fluorescent dyes linked to the nucleobases, but their enzymatic incorporation becomes increasingly limited with large fractions of labeled dNTP replacements. Current solutions for most sequencing technologies are adapting stepwise additions of base-labeled nucleotides, followed by chemical or photochemical removal of the label, resulting in reduced sequencing speeds as additional washing and cleavage steps have to be performed (Ju et al. 2006; Korlach et al. 2008; Mitra et al. 2003).

In SMRT sequencing, an alternative approach is applied that attaches the fluorescence label onto the phosphate chain instead of the base. In this case, as the DNA polymerase induces the cleavage of the α-β-phosphoryl bond in dNTP during DNA synthesis, a pyrophosphate with the attached fluorescent label is released, leaving a natural unmodified nucleotide in the newly synthesized DNA strand. Linking a fluorescent dye directly onto the phosphate in dNTP introduces steric hindrance as a potential cause of DNA polymerase inhibition; however, an extension of the triphosphate moiety to four or five phosphates was reported to increase incorporation efficiency (Kumar et al. 2005). The form of the labeled nucleotides used in SMRT sequencing is that fluorescent dye is conjugated to an aliphatic linker that separates the nucleotide and the fluorophore thus allowing larger spatial separation, and then built onto pyrophosphate moiety. By using terminal phosphate-labeled nucleotides, the "cleavage and washing" scenario is avoided and therefore realizes sequencing in real time and shortens time to result dramatically. The overview of SMRT sequencing is shown in Figure I-18.

Unlike NGS and the other two third generation sequencing platforms, SMRT sequencing is capable to read up to 10,000 bases with an average of 1,000 bases long reads. High processivity is achieved by using Φ29 DNA polymerase which is also capable of strand displacement DNA synthesis, enabling the use of double strand DNA as template. Φ29 DNA polymerase has also been currently widely used in whole genome amplification approaches (Dean et al. 2002; Silander and Saarela 2008). A wild type of Φ29 DNA polymerase was modified to have improved performances in sequencing. The mutant has reduced 3'-5' exonuclease activity but maintains the identically polymerization properties as the wild type (Korlach et al. 2008).

The SMRT sequencing platform provides three read types: (1) standard sequencing in which a long inserts library is made so that DNA polymerase can synthesize along a single strand; (2) circular consensus sequencing (CCS) in which insert size is short and double strand template is ligated to a pair of hairpin-like adapters so that both the forward and reverse strand can be read for several times each (Figure I-18); (3) strobe sequencing in which requires very long insert size, the laser light in the instrument is alternated between on and off during sequencing step so that on-periods generate the sequencing reads and off-periods determine the length of the space in between.

Fluorescence pulses in SMRT sequencing are not only characterized by their emission spectra but also by their duration and by the interval between successive pulses, from which two parameters are obtained: pulse width (PW) and interpulse duration(IPD), reflecting the kinetics of the polymerase while the sequencing is in process. PW is a function of all kinetics steps after nucleotide binding and up to fluorophore release, whereas IPD is determined by the kinetics of nucleotide binding and polymerase translocation. Eid et al also reported that the IPD was strongly affected by the DNA template whereas the PW was governed by local chemical processes in the active site so that PW showed only moderate variability with sequence context (Eid et al. 2009). A SMRT cell contains approximately 75000 ZMW in which about one third contain a single DNA polymerase with optimized loading. The DNA synthesis rate is about 2~4 bases per seconds and therefore a single SMRT sequencing run takes only a few hours. The current error rate of 15 % is significantly higher than with other sequencing techniques, which a proeminence of deletions, followed by insertions rates. The deletions probably stem from incorporation events or intervals that are too short to be reliably detected while the insertions may be caused by dissociation of a cognate nucleotide from the active site before phosphodiester bond formation resulting in the duplication of a pulse. Although the current error rate is high, the erroneous position happens stochastically during

sequencing. So the error rate can be diminished by CCS in which both strands are read several times. In an approach followed by Travers and his colleagues (Travers et al. 2010), first a double strand template, with both ends ligated with a hairpin-like adaptor was used to constructed the library called SMRTbell, thus sequencing by CCS read type as described above. With an insert length of 250bp, theoretically, an expected phred-style quality value could reach 30 which is sufficient for SNP detection. The accuracy is positively related with the sequencing depth, it is reported that with 15-fold average coverage, the median accuracy can achieve 99.3% (Eid et al. 2009).

SMRT sequencing has a fascinating utility in detecting DNA methylation (Flusberg et al. 2010) and damaged DNA bases (Clark et al. 2011). Both studies are based on the principle that the kinetics of DNA polymerase is influenced by DNA sequence context. Compared with bisulfite conversion combined with massively parallel sequencing, SMRT sequencing provides opportunities for the direct detection of single DNA molecule methylations without bisulfite conversion which simplifies the sample preparation and reduces the complexity in post-sequencing analysis. Furthermore, different modifications such as N6-methyladenosine, 5-methylcytosine and 5-hydroxymethylcytosine influence the kinetics of DNA polymerase in different patterns, the assignment and classification of the modifications can therefore be inferred from the metrics of PW and IPD. The discrimination between cytosine, 5-methylcytosine and 5-hydroxymethylcytosine cannot be accomplished with bisulfite sequencing. The Pacific Bioscience company is still refining this technique to make *de novo* methylation profiling become possible.

2.5.7 De novo assembly for TGS

The TGS technologies described above devote many efforts to reducing the sequencing biases caused by PCR amplification to generate template clusters, to produce long sequencing reads, to shorten the run times and to reduce the instrument cost by avoiding optical system in base identification. But in the library preparation step, all the TGS technologies still use the *in vitro* library preparation strategies as for NGS (or second generation sequencing) so that the size of the inserts is still limited to 20kb which still makes the large eukaryotic genome difficult assemble into ultracontig or superscaffold. The final solution may still need mapping-based strategies to order and assign the scaffolds onto chromosomes.

3. Avian Genome Structure

It is believed that avian species could have existed at least since the late Triassic period which is about 200 million years ago since discovery of two nearly complete fossil skeletons of *Protoavis* which pre-date the Jurassic Archaeopteryx by some 50 million years. Mitochondrial analysis suggested that the common ancestor of birds and mammals diverged 310 million years ago while the common ancestor of birds and crocodilians diverged 210-250 million year ago (Burt et al. 1999; Griffin et al. 2007; Muller and Reisz 2005). The evolutionary relationships among major avian groups are contentious although well studied (Chojnowski et al. 2008; Ericson et al. 2006). But there are two nodes at the base of the avian tree that are supported by both morphological and molecular phylogenetic studies (Chubb 2004; Groth and Barrowclough 1999; Hackett et al. 2008). The first divides into the Paleognathae (ratites and tinamous) and Neognathae (all other birds), and the second splits the neognaths between the Galloanserae (Galliformes and Anseriformes) and Neoaves (other neognaths). According to the data from Timetree website (http://www.timetree.org/), the mean divergence between Galliformes and Anseriformes is about 81.2 million years. Although many bird species have diverged tens of millions years or even longer, avian species possess highly conserved karyotype and synteny (Nanda et al. 2011; Shibusawa et al. 2004).

Most avian species contain about 40 pair of chromosomes except some notable extremes like the stone curlew and kingfisher, with 20 and 66 pairs of chromosomes, respectively (Burt 2002). Of 40 pairs chromosomes, seven or eight pairs are the largest chomosomes, the macrochromosome which are 3μm ~ 6μm in length; the remainings are 0.5μm ~2.5μm in length and named as microchromosomes (Rodionov 1996). Interestingly, in Accipitridae, the total number of chromosomes is about 70 but they only have 3 to 5 pairs of michromosomes (Bed'Hom et al. 2003). The organization of their karyotype is really different

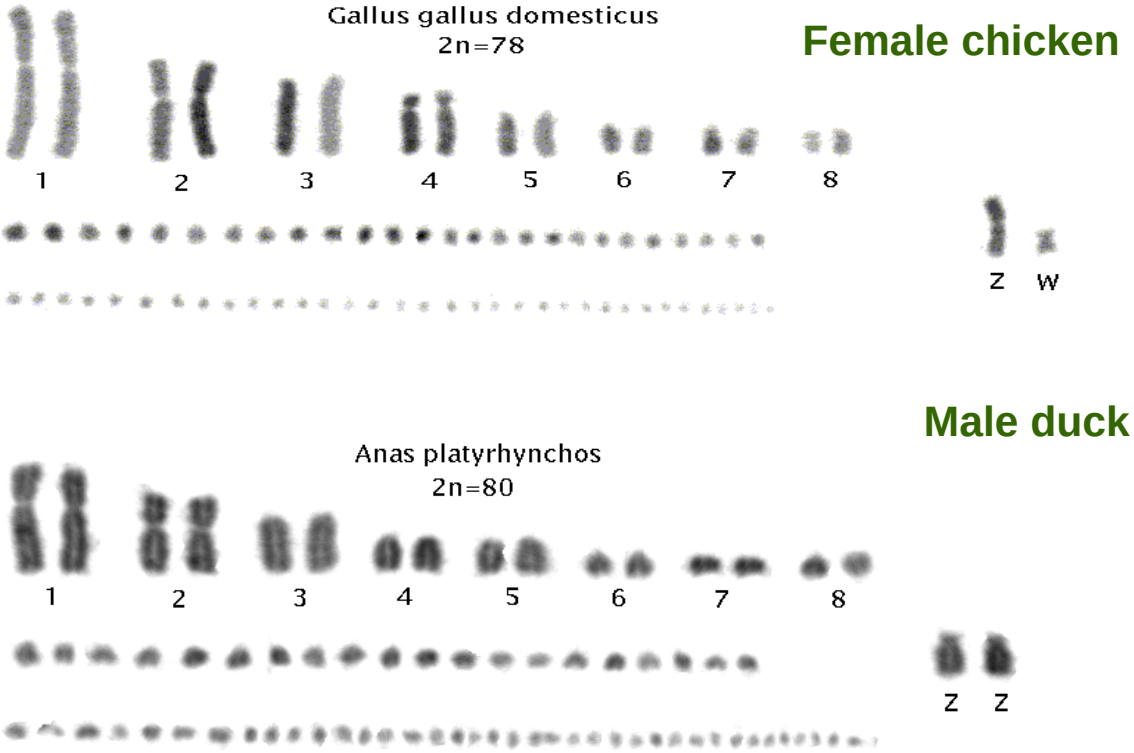


Figure I-19: Karyotype of a female chicken and a male duck. (V. Fillon personnal communication).

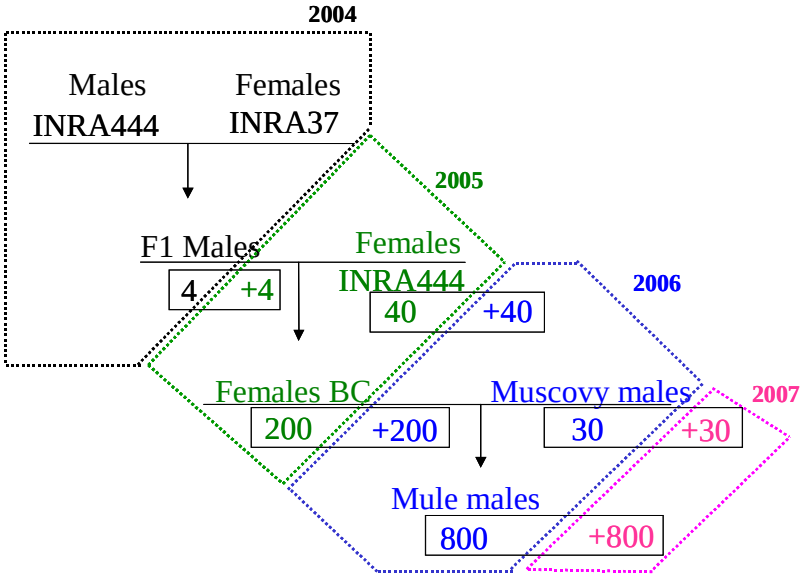


Figure I-20: The French genetic mapping and QTL resource family. The family structure, a backcross design, is designed to detect traits influencing the overfeeding efficiency and liver lipid metabolism. The originality of the design is that the phenotypes of the BC daughters are estimated through the performances of their Mule duck offspring. The trait of interest is thus the genetic capacity of the BC female Beijing ducks, to improve the performances of their sons. (C.Marie-Etancelin personnal communication)

than the classical bird karyotype. In birds, the nomenclature of sexual chromosomes is different from mammals which are named as Z and W rather than X and Y. In contrast to mammals, the females are heterogametic in which karyotype is ZW and males are homogametic whose karyotype is ZZ in birds. Moreover, comparative genomics showed that ZW chromosomes are not syntenic to mammalian XY but mostly syntenic to HSA5 and HSA9 (Fridolfsson et al. 1998; Nanda et al. 1999; Stiglec et al. 2007).

Of all the avian species, chicken is the most studied. It has a karyotype composed of 39 pair chromosomes in diploid cells, in which 30 are small to tiny microchromosomes and a pair of Z and W sex chromosomes (Burt 2002). The ancestral karyotype of the birds appears similar to the chicken one, with GGA1, 2, 3, 4q, 5, 6, 7, 8, 9, 4p and Z representing the ancestral avian chromosomes 1-10 + Z (Griffin et al. 2007). Sequencing comparison between chicken and human reveals that all the chromosomes show extensive interchromosomal rearrangements except for HSA4 and GGA4q, so it is speculated that GGA4q is the most ancient chromosome and appeared at least 310 million years ago before the divergence of birds and mammals. The rest of the ancestral avian chromosomes appeared at least 210 million years ago (Chowdhary and Raudsepp 2000; Griffin et al. 2007).

The karyotype of duck (2n=80) is very similar to that of chicken (2n=78) except for one known interchromosomal rearrangement, with GGA4 (chicken chromosome) corresponding to APL4 and APL10 (duck chromosomes), explaining the difference in chromosome numbers in the karyotypes (Denjean et al. 1997; Fillon et al. 2007; Ladjali-Mohammededi et al. 1999; Skinner et al. 2009) (Figure I-19). There are no more interchromosomal rearrangements known to date between chicken and duck, and the published comparative cytogenetic maps only detected a few intrachromosomal rearrangements on some macrochromosomes. Due to due to the low resolution of the fluorescent *in situ* hybridization (FISH) techniques used, no intrachromosomal rearrangement

have been observed to date on microchromosomes. High synteny conservation is observed not only between chicken and duck (Fillon et al. 2007; Skinner et al. 2009) but also among other birds such as between chicken and turkey (Griffin et al. 2008; Zhang et al.), chicken and zebra finch (Volker et al.; Warren et al. 2010), chicken and quail (Kayang et al. 2006; Sasazaki et al. 2006).

Compared to mammals, most birds have a genome which is approximately three times smaller although the content in genes is expected to be very similar. The smallest bird genome is about 0.91pg per haploid genome for the Black-chinned hummingbird (*Archilochus alexandri*) and the largest is 2.16pg for the ostrich (*Struthio camelus*) (http://www.genomesize.com/statistics.php?stats=birds). One hypothesis is that the smaller genome could be related to the energy conservation requirements associated with the evolution of flight (Hughes and Hughes 1995). However, a study made by Organ et al has shown that the small genome of birds originate deep within the dinosaurian roots of modern birds, long before the origin of flight, perhaps as a means of accommodating other metabolic needs (Organ et al. 2007). However, this hypothesis may be supported by data on the bat genome, as a bat is a mammal which can fly, and also by the ostrich genome (2.16pg for ostrich versus mean value of 1.38pg for birds). Indeed, the mean genome size for mammal is approximately 3.37pg whereas a bat species (*Miniopterus schreibersi*) has a genome whose estimated size is only 1.77pg. The gene counts are similar between birds and human while the genome size is significantly smaller in birds. This can be explained by a much lower repetitive content of the genome and smaller introns in birds than human as first revealed by the chicken genome sequences (Consortium 2004).

3.1 Sex Chromosome

The sex chromosomes, either XY or ZW, often show different size, structure and gene content (Otto et al. 2011). In fact, avian Z and W sex chromosomes share some common features with mammalian X and Y sex chromosomes: (1) all of them contain many repeat sequences which is a major reason responsible for the difficulties in sequencing and assembly; (2) both the X and Z chromosomes are extremely conserved; (3) both Y and W are degenerated and highly heteromatic in most species and (4) both XY and ZW only pair and recombine at the tips known as the pseudoautosomal regions during meiosis.

3.1.1 Evolution of sex chromosomes

It is believed that sex chromosomes originate from an ordinary autosome pair via the acquisition of a dominant sex determination gene (Charlesworth and Charlesworth 2005; Malone and Oliver 2008). The Human sequencing project showed that HSAX is largely euchromatic but that 56% of the euchromatic regions are interspersed repeats, and the GC content is 39%, which is lower than that of genome average (41%). The evolution of the sex chromosomes with the shrinkage of the Y and W has been described by the principle of Muller's ratchet (Muller 1964). Briefly, if there is no recombination as it is the case for most of the length of the Y and W chromosomes, when linkage groups carrying the fewest number of deleterious mutations are lost in the species population, there is no way back and the genetic load increases, leading to a gradual deterioration of the chromosome. This hypothesis is widely accepted for the evolution of sex chromosome due to their unique characteristics. The hypothesis explaining the absence of recombination for the Y and W chromosomes is the necessity of conserving intact the sex determining factors, which are under strong constraint. Thus in the absence of recombination, some deleterious mutation would accumulate and those regions would be gradually eliminated during the evolution. Generally speaking, the way that sex chromosomes evolve is very similar between birds and mammals.

3.1.2 Dosage compensation

Dosage compensation is an epigenetic mechanism that normalizes the expression of genes on the sex chromosomes, between the individuals having two (XX or ZZ) or one copy (XY or ZW) (Conrad and Akhtar 2012). Different organisms use different strategies to solve the balance of gene expression for the X and/or Z chromosomes. Studies in model organisms show that there are mainly two different strategies for dosage compensation: one is to double the gene expression level on X and/or Z chromosomes in the heterogametic sex like in the fruit fly (Gorman and Baker 1994; Prestel et al. 2010) and another involves inactivation of one of the X and/or Z chromosomes in the homogametic sex such as in human and mouse (Brown et al. 1991; Heard and Disteche 2006).

In birds, however, there is a debate about the existence of dosage compensation. In the last century, it was widely accepted that dosage compensation did not exist in birds (Baverstock et al. 1982), based on the observation of the absence of a Barr body or a late replicating Z chromosome in male avian nuclei (Schmid et al. 1989). However, there are some recent reports suggesting that dosage compensation may exist in birds (Kuroiwa et al. 2002; McQueen et al. 2001). The contradictory conclusions on dosage compensation in birds are probably due to the limited number of genes investigated, as some may escape from inactivation if it does exist. Nevertheless, several other findings fueled the idea that dosage compensation is weak in birds. For instance, sexual dimorphism approaches in gene expression found a disproportionate number of Z genes among male-biased genes and genes with male-specific expression were disproportionately Z-linked rather than autosomal in gene expression databases (Agate et al. 2004; Agate et al. 2003; Arnold et al. 2008; Chen et al. 2005; Scholz et al. 2006; Storchova and Divina 2006). The mechanism of dosage compensation in birds, if birds really do have one, is certainly different from the one in mammals in which one copy of the X chromosome is inactivated so that the X-linked genes

are transcribed from only one activated X chromosome. In birds, a biallelic expression pattern is observed for Z genes, suggesting that one of Z chromosome is not inactivated in the same manner (Kuroda et al. 2001; Kuroiwa et al. 2002).

X chromosome inactivation in mammals is triggered by a non-coding specific transcript named Xist which is devoid of any significant ORF and expressed from the inactive X chromosome in somatic cells at the X inactivation center (Augui et al. 2011). Xist can coat the chromosome from which it is expressed (Clemson et al. 1996) and a complex pathway is employed to cause hypermethylation and heterochromatization of the entire X chromosome., leading to its inactivation except for a few genes that escape. In chicken, a region located on the p arm of the Z chromosome shows a lower male:female (M:F) expression ratio than the rest of the chromosome, suggesting a regional dosage compensation (Melamed and Arnold 2007). Inside this region, there is a Z non-coding RNA (ncRNA) showing a female-specific expression pattern. The ncRNA is only expressed in females at the Male HyperMethylated (MHM) locus, probably because the DNA at MHM locus is hypermethylated and transcriptionally silenced in ZZ males (Teranishi et al. 2001). However, in zebra finch no low (M:F) ratio was observed making dosage compensation and Z inactivation more complicated (Warren et al. 2010).

To sum up genetic studies on sex chromosomes in birds so far show that dosage compensation is not obvious as compared to mammals and fruit fly. However it is evident that dosage compensation does happen in birds, at least in some species in a region-wise manner.

3.2 Sequenced Avian Genomes

3.2.1 Chicken Genome

Chicken is the first sequenced livestock and bird species. The first draft genome was obtained by Sanger sequencing from an inbred female red jungle fowl to minimize

heterozygosity and provide sequence for both sex chromosomes in 2004. The assembly was generated from 6.6-fold coverage whole genome shotgun reads from plasmid, fosmid and BAC-end read pairs. The scaffolding and chromosomal assignment were accomplished by combining a BAC-based physical map and a genetic map, and thereafter the final assembly was improved by including expression sequence tag (EST) and mRNA data. The final assembly was 1.05 Gb in which 933Mb were assigned to specific chromosomes and the remaining were placed on a virtual chromosome, chrUN (chromosome Unknown) (Consortium 2004).

Several insights have been yielded from chicken sequencing: (1) the chicken genome is almost one third of a typical mammalian genome in size, mainly due to the repeat content which occupies around 15% of the assembled chicken genome in contrast to around 50% in mammals; (2) GC content, CpG island, recombination rate and synonymous substitution rate are negatively correlated with chromosome size; (3) there is a paucity of retroposed pseudogenes in the chicken genome and (4) alignment of the chicken and human genome identifies at least 70 Mb of sequence that are highly conserved and thus have a high probability of being functional in both species.

a. Genome Content of Chicken Genome

An evidence-based system and two comparative gene prediction methods together predicted a common set of 106,749 protein-coding exons which may represent around 20,000 to 23,000 protein-coding genes. Alignments of chicken and human orthologous protein-coding genes demonstrate the expected pattern of sequence conservation, with the highest sequence similarities in protein-coding exons and in introns. Moreover, the alignment of orthologous coding regions often did not extend in 5’ to the previously annotated human

protein start codon, indicating that an internal ATG codon could be the true translation start site for approximately 2,000 human genes, thus improving the annotation of the human genes.

In the chicken genome, only 51 retrotransposed pseudogenes were found in contrast to more than 15,000 copies in rat and human genome (Gibbs et al. 2004; Torrents et al. 2003). Among those 51 gene duplicates 36 are clearly pseudogenes, and there is no clear bias towards either particular gene families or chromosomal locations (Consortium 2004). The low number of pseudogenes might be linked to resistance mechanisms towards the invasion of repetitive elements.

Insights into the repetitive content of the the chicken genome show a dominance of transposable elements (TEs). The most abundant TEs are a family of non-long terminal repeat (LTR) retrotransposon called CR1 (Chicken Repeat 1). CR1 resembles mammalian L1 elements whose full length is estimated between 6~7 kb and having a GC-rich internal promoter region, followed by two open reading frames (ORF) with the second ORF encoding a reverse transcriptase (Mathias et al. 1991). The full length of a CR1 is 4.5kb, but more than 99% of the CR1 copies in the genome are truncated at the 5' end and most CR1 elements in chicken are less than 500 bp long (Wicker et al. 2005). The CR1 elements are mainly divided into six large subfamilies designated A-F, in which B, C, D and F subfamilies probably have descended from four different progenitors whereas A and E subfamilies may have been spawned from the ancestor of those four different progenitors or from a distinct progenitor (Vandergon and Reitman 1994). Although CR1 elements resemble mammalian L1 in some aspects like their abundance or general structure, the consequences of retrotransposition and the evolution mechanisms are different. CR1 do not create target site duplication (TSD) which is a typical byproduct in mammalian retrotransposition (Martin et al. 2005), and the evolution of CR1 in birds suggesst that widely divergent elements have been active in parallel whereas in mammals a single lineage of L1 elements has been dominant (Adey et al. 1994; 2004; Smit

et al. 1995). Besides, the most consequential difference in the structure of CR1 and L1 is in their 3'end: the 3'UTR of L1 elements are divergent from their ancestors except for the polyadenylated tails whereas the CR1 3'UTR are remarkably conserved between all derived subfamilies and end with microsatellite repeats in all chicken CR1 subfamilies (Haas et al. 2001; Smit et al. 1995).

Another striking discovery is that in chicken there is a paucity of short interspersed element (SINE). SINEs are small, non-autonomous retrotransposons derived from structural RNA having an internal polymerase III promoter. Generally, the retrotranspositon of SINEs relies on the replication machinery of the autonomous retrotransposons (Kramerov and Vassetzky 2005). In contrast to mammals, SINEs occupy 7% of the genome in the rat (Gibbs et al. 2004), 8% in the mouse (Waterston et al. 2002) and 13.64 % in the human genomes (Lander et al. 2001), whereas in chicken there is not a single SINE, although there are about 10000 faint matches in the chicken genome to MIR and MIR3 (the SINEs associated with L2 and L3 respectively).

Furthermore, the chicken genome sequence provides clear evidence that macrochromosomes and microchromosomes have distinct genomic features. Previous studies suggested that microchomosomes are CpG-rich and gene-rich, reflecting high transcriptionally activities (Andreozzi et al. 2001; Grutzner et al. 2001; Habermann et al. 2001; McQueen et al. 1996; McQueen et al. 1998; Ponce de Leon et al. 1992; Schmid et al. 1989; Smith et al. 2000). The macrochromosomes represent two thirds of the genome but only just half of the genes. Compared to microchromosomes, machromosomes contain more repetitive elements, a lower gene density and also exhibit a lower rate of synonymous substitutions, although they have the same rate of non-synonymous substitutions. Alignment to the genetic map also showed that microchromosome have higher recombination rate than macrochromosomes (median value of 6.4cM/Mb and 2.8cM/Mb, respectively).

3.2.2 Zebra Finch genome

The zebra finch, an important model organism in neuroscience, is the second bird sequenced (Clayton et al. 2009; Doupe and Kuhl 1999; Kuhl 2003). Zebra finch communicates through learned vocalizations, an ability otherwise documented for only in humans and a few other animals and absent in chicken (Jarvis 2004; Warren et al. 2010). Unlike chicken, zebra finch belongs to the largest orders in the Ave class, Passeriformes (Hackett et al. 2008). The overall structure of the genome is similar between chicken and zebra finch; however, they differ by many intrachromosomal rearrangements, lineage-specific gene family expansions and repeat content composition.

The zebra finch genome was assembled by Sanger pair-end sequencing of plasmid, fosmid and BAC libraries from a single male individual. The initial assembly was based on 6X coverage whole genome shotgun reads and then improved with 35 finished BAC clones which led to a 1.2 Gb draft genome. The N50 contig size is 36kb and 39kb for chicken and zebra finch respectively. The length of N50 scaffold is 7 Mb and 9 Mb for chicken and zebra finch respectively.

The zebra finch genome contains half of the chicken's CR1 content but three times more retrovirus-derived LTR than chicken. More surprisingly, in the zebra finch genome, a low copy number of SINEs are found which are absent in the chicken genome. Expressed sequence tag (EST) analysis shows that mobile elements are present in 4% of the transcripts expressed in the zebra finch brain and some of them are regulated by song exposure (Warren et al. 2010).

3.2.3 Turkey genome

A female turkey was sequenced using multiple sequencing platforms. The sequencing reads for the genome assembly were produced solely from the Roche 454 GS-FLX and the

Illumina Genome Analyzer *II* platforms and additionally, 400 000 BAC ends were sequenced by Sanger sequencing for linking scaffolds and for chromosome assignment. The Illumina platform was used to produce single and pair-end reads from short insert libraries (180 ~200 bp) while the 454 platform generated sequencing reads from long insert libraries (3kb and 20kb). The draft assembly spans 1.038 Gb in which the N50 size of contig and scaffold is 12.6kb and 1.5Mb respectively. The repeat content and gene content are very similar between chicken and turkey, but slight differences are observed. Compared to chicken, turkey has a lower repeat content, with 6.94% of the assembled draft genome. In term of gene content, the overall gene content is similar except some new families with unknown functions.

3.3 Avian comparative Genomics

Hitherto there are three sequenced avian species published. Genome comparison provides further evidence of the high level of karyotype and chromosome synteny conservation in birds. Only a few cases of interchromosomal rearrangements are reported, most of which caused by fission or fusion events explaining the previously observed karyotype differences such as the number of chromosomes or of chromosome arms. Intrachromosomal rearrangements are more frequent and thus are speculated to play an important role in speciation.

The genetic maps, physical maps and genome sequences have revealed highly conserved synteny and a few chromosomal rearrangements among chicken, zebra finch and turkey. A diploid genome of zebra finch contains 40 pairs of chromosomes whereas chicken has 39 pairs. The sequence of zebra finch genome and FISH experiments confirm the high degree of almost one-to-one homology between chicken and zebra finch that had been suspected from genetic mapping results (Stapley et al. 2008; Warren et al. 2010). The genetic map of zebra finch has confirmed two interchromosomal rearrangements documented in 2004,

in which GGA1 corresponds to two large zebra finch chromosome TGU1 and TGU1A and GGA4 corresponds to macrochromsome TGU4 and microchromosome TGU4A (Derjusheva et al. 2004). More intrachromosomal rearrangements than expected have been revealed by comparing zebra finch genetic map with the chicken genome, suggesting that the gene order is not highly conserved between Passeriformes and Galliformes after they diverged from their common ancestor about 100 MYA ago (http://www.timetree.org) (Pereira and Baker 2006). The intrachromosomal rearrangements when compared to chicken are not only inversions, but also involve translocation and more complex rearrangements. Additionally, genome sequencing also reveals that the major histocompatibility complex (MHC) is dispersed across several chromosomes, whereas is present at two loci, but only on a single microchromosome in chicken (Consortium 2004; Warren et al. 2010).

Previous studies have already shown that there is a high degree of synteny and karyotype conservation between chicken and turkey despite 20 ~ 40 million years divergence (Dimcheff et al. 2002; van Tuinen and Dyke 2004). There are two interchromosomal rearrangements between chicken and turkey due to translocations. One event is probably due to a fission in the turkey lineage and as a result, GGA2 corresponds to MGA3 (turkey chromosome) and MGA6. Another event is a fusion in the chicken lineage in which GGA4 corresponds to MGA4 and MGA9 (Dalloul et al. 2010; Griffin et al. 2008; Reed et al. 2005; Reed et al. 2007). There are 20~27 major rearrangements predicted between chicken and turkey and all the cytogenetic experiments have shown that most of the intrachromosomal rearrangements are pericentric inversions resulting in the turkey chromosomes being prone to be telocentric (Zhang et al. 2012). It is suggested that there might be a fusion event in the turkey lineage involving two ancestral microchromosomes fused into one larger microchromosome, but this is not evident in cytogenetic mapping. Moreover, there are still

~10 microchromosomes missing in all the sequenced birds, so that the sequencing data is not available to support this hypothesis.

On the contrary, in mammalians comparative genomics reveals both extensive interchromosomal and intrachromosomal rearrangements (Gibbs et al. 2004; Gregory et al. 2002; Zhao et al. 2004). For instance, although mouse and rat have diverged from their common ancestor only 12 ~ 24 MYA ago (Adkins et al. 2001; Springer et al. 2003), mouse has one extra chromosome pair. Except for a few exceptions such as MMU4, MMU9 and MMUX (mouse chromosomes) having a one-to-one well-conserved synteny to RNO5, RNO8 and RNOX (rat chromosomes) respectively, all the other chromosomes demonstrate inter-chromosomal rearrangements during evolution. The intrachromosomal rearrangements between human and chimpanzee which diverged around 6MYA ago (Chen and Li 2001), also outnumber those between chicken and turkey despite a longer evolution time for the latter pair.

Although chromosomal rearrangements both between the mouse and rat or the human and chimpanzee pairs of species may be extreme examples, the average number of chromosomal rearrangements in mammals is relatively higher than in birds. Burt *et al* have revealed that the organization of the human genome is closer to that of the chicken genome while comparing human, mouse and chicken (Burt et al. 1999). One possible explanation is that both human and mouse contain much more transposable elements and repeats, so that the rearrangements by illegitimate recombination are more common; the rates for human and mouse lineages are 0.58 and 1.14 rearrangements per MYA. It is proposed that transposable element (TE) may be the driving force for chromosome evolution. This hypothesis has emerged from the analysis of large scale rearrangements by comparing different sequenced species, in which an enrichment of TE has been observed at the breakpoints(Eichler and Sankoff 2003). Comparison of human and mouse(Dehal et al. 2001) and of the three sequenced birds (Skinner and Griffin 2011), supports the fact that the breakpoint regions

where the rearrangements happen are significantly enriched with repeats. Thus it is postulated that highly conserved karyotypes and syntenies result from the lower repeat content in bird genomes.

4. Current status of duck genomics

4.1 Duck genetic map

The first duck genetic linkage map was developed by using a cross between two extreme Beijing duck lines by Huang et al in 2006 (Huang et al. 2006). These two lines were selected for high body weight at 42 days of age or high egg production at 360 days of age, and an experimental population with a total of 224 G2 individuals was created. Linkage analysis of 155 polymorphic microsatellite markers performed on this population produced the first duck genetic map containing 19 linkage groups. Out of 155 microsatellite markers genotyped, 115 are placed on the genetic map. The sex averaged map spans 1353.3cM, with an average interval distance of 15.04 cM. The male map covers 1,415cM whereas the female map covers 1387.6cM. The flanking sequences of 155 genotyped microsatellite markers were aligned on the chicken genome by BLASTn and 49 corresponding ortholog sequences were found. Specific PCR primers were designed based on the corresponding orthologs and used to select 28 chicken BAC clones which were then used to integrate genetic and cytogenetic map by FISH. Eleven out of 19 linkage groups were thus assigned to 10 duck chromosomes.

The first QTL detection on carcass and meat quality traits was carried out by Huang *et al* in 2007 (Huang et al. 2007b), based on the Chinese resource family used for building the map (Huang et al. 2006). With the 95 microsatellite markers tested, eight genome-wide significant QTL for crop weight, skin fat, liver weight, neck, shanks, wings and drip loss were detect on linkage groups CAU4 and CAU6; one genome-wide suggestive QTL and one chromosome wide QTL affecting breast weight were detected on linkage groups CAU1 and CAU4 respectively. Fifteen chromosome-wide suggestive QTL influencing weight of abdominal fat, breast, crop, heart, carcass, thighs, liver, shanks, gizzard, fat thickness in tail,

drip loss and cooking loss were mapped to linkage groups of CAU2, CAU4, CAU5, CAU6,CAU7, CAU10 and CAU13. A second QTL detection on traits influencing body weights and conformation were perform in the Chinese resource family (Huang et al. 2007a). Six genome-wide suggestive QTL for three body weight traits and two body conformation traits were identified on the linkage groups of CAU1, CAU2, CAU6 and CAU12. Chromosome-wide significant QTL affecting body weight traits and one conformation trait were found on CAU4 and CAU10. Besides, 12 chromosome-wide suggestive QTL for 6 body weight traits and 4 body conformation traits were located on seven different linkage groups. Moreover, the QTL on CAU6 at 21cM and 73cM jointly influenced shank girth and could explain 10.6% of phenotypic variations.

A second duck genetic map has been constructed from a resource family in France (Marie-Etancelin *et al*. *in prep*). This resource family was designed to detect and map single and pleiotropic QTL segregating in the Common duck having an influence on the expression of traits in their overfed mule duck offspring. To this end, a Common duck back cross (BC) design has been generated by crossing Kaiya ducks (I444) which are from a light strain and heavy Beijing ducks (I37) (Figure I-20). The two lines differ notably in the bodyweight and overfeeding ability of their mule progeny. The BC females were mated to Muscovy drakes, and their mule duck progenies were measured for growth, metabolism during growth and overfeeding period, overfeeding ability, breast meat and fatty liver qualities. The phenotypic value of each BC female was estimated for each trait by assigning the mean value of its offspring's phenotype values, taking into account the variance, which depends on the number of sons measured per BC and the heritability of the trait considered.

The genetic map used for QTL detection has 91 microsatellite markers aggregated into 16 linkage groups (LG), covering a total of 778 cM. Twenty-two QTL were found significant

at the 1% chromosome-wise threshold level, using the single trait detection option of the QTLMap software. Most QTL were detected for breast meat and fatty liver qualities: QTL for meat pH 20 minutes *post mortem* were mapped on LG4 (at 1% genome-wide level) and QTL for meat lipid content and cooking losses were found both on LG2a. For the fatty liver weight and composition in protein and lipid, QTL were mainly detected on LG2c and LG9 and multiple traits analyses highlighted pleiotropic effects of QTL in these chromosome regions. Apart for the strong QTL on chromosome Z for plasma triglyceride content at the end of the overfeeding period detected in single trait analysis, all metabolic traits QTL were revealed with the multi-traits approach: QTL on LG14 and LG21 affected the plasma cholesterol and triglyceride contents whereas QTL on LG2a seemed to impact glycaemia and the basal plasma corticosterone content (C. Marie-Etancelin *et al, in prep*).

4.2 BAC library & Fosmid library

Moon and Magor constructed a duck fosmid library for comparative genomic analysis in 2004 (Moon and Magor 2004). Before this period, others had tried to construct a BAC library for duck but failed due to problems with recombination, insertions and deletions (Moon and Magor 2004). A male Beijing duck was the DNA source and known to be heterozygous for MHC class I genes. This individual was chosen for two purposes: on one hand, he was the principal breeding male of the University of Alberta duck colony; thus, his haplotypes should be found in many offspring and available for future studies. On the other hand, the cDNA library constructed from his spleen will allow the comparison of expressed genes to those present in the genome within one individual. The final fosmid library consists of 124,488 clones and is estimated to have genome coverage of 4.7X with an average insert size of 38kb.

A duck BAC library deriving from an inbred Beijing duck, which is the individual that was subjected to whole genome sequencing afterwards, has been constructed in 2006 by partial digestion with *HindIII* restriction enzyme and then ligated to the pIndig-5 vector (Yuan et al. 2006). The whole library comprises 84,480 clones representing nine-fold physical coverage of the duck genome. The estimated average insert size of this library is close to 118kb (Yuan et al. 2006).

4.3 SNP Detection

Kraus *et al* have reported a genome wide SNP discovery from nine wild mallard ducks collected from three locations in Europe (Kraus et al. 2011). More than 122,000 SNPs were identified within this sample by sequencing a reduced representation library at a depth of 16 X. All the sequencing reads were then mapped to the duck draft assembly thus identifying 62,000 additional SNP. Altogether more than 184,000 SNP were identified from this study in which almost 150,000 have the characteristics required for subsequent genotyping. Among those high quality 150,000 SNP, approximately 101,000 SNP were detected within wild mallard sequences and the rest were detected between wild mallard and domesticated duck. Within the dataset of 101,000 SNP, they found a subset of ~20,000 shared between wild mallard and domesticated duck, suggesting a low genetic divergence (Kraus et al. 2011).

One run of sequencing including F1 animals of the resource population used for the construction of the genetic map and for QTL research is scheduled at INRA (Frédérique Pitel and Alain Vignal, INRA, France). This approach should produce less SNP per kb of chromosome sequence due to the limited number of animals coming from the cross of two

domestic lines, but these should be the best choice for genotyping the French QTL resource family, due to the choice of animals to be sequenced.

4.4 EST data

Most of the efforts are devoted towards the production of Expressed Sequence Tags (EST), which will be subsequently used to annotate the genome and to design chips for transcriptome analyses. Eight runs of sequencing with a Roche 454 have been performed and are under analysis (Dave Burt, UK, personnal communication). These include tissues involved in the response to the Influenza virus, mainly spleen, lung and intestine. Control and challenged birds are included in the analysis. Another analysis is performed at INRA (Frédérique Pitel, Christian Diot and Alain Vignal, INRA, France) and consists of 2 runs of Roche 454 sequencing of liver, muscle and brain tissues from both the common duck and the Muscovy duck. All the sequencing data were assembled into 64,946 EST contigs and used to annotate the duck genome scaffolds sequenced at BGI (Huang et al, in prep).

4.5 Duck genome sequencing

A 10-week-old female Beijing duck from the Golden Star Duck Production in China has been sequenced by the BGI (Beijing Genomics Institute) using similar methods as those used for the sequencing of the giant panda genome (Huang *et al, in prep*). The genome analysis is mostly finished. In total they generated 77 Gb of paired-end and mate-pair reads representing a 64X physical coverage of the genome with an average read length of 50bp. The assembly is composed of 78,487 scaffolds covering 1.1 Gb. The N50 contig size is 26kb and the N50 scaffold is 1.2Mb.

All the EST data were collected and combined for improving the assembly and gene prediction, which resulted in 15,065 protein-coding genes in duck. Genscan and Augustus, trained on human data, were used to predict duck genes, giving a prediction of 32,383 and 22,739 protein-coding genes respectively (Huang *et al, in prep*). After integrating all gene sources, a reference gene set was created containg 15,634 protein coding genes, 249 pseudogenes and 567 ncRNAs occupying 2.3% of the duck genome (http://pre.ensembl.org/Anas_platyrhynchos/Info/Index).

The whole genome sequencing predicted almost 2.8 million SNPs from which the estimated the heterozygosity values of the duck genome were estimated to be 2.61×10^{-3} for the autosomes in general and 2.08×10^{-3} for the coding regions(Huang *et al, in prep*) . The transcriptome sequencing data from a cherry valley duck was mapped to the draft assembly, increasing the total SNP number to more than 2.95 million. Therefore, on average there are 2.76 SNPs per kb in the duck genome when comparing two random complements. The fraction of SNP in the intergenic, intronic and exonic regions is 63%, 34.3% and 2.7% respectively(Huang *et al, in prep*) .

It is estimated that segmental duplications (SD) represent 1% of the duck assembly, which is similar to chicken but significantly less than mammals for which SD represent 3.1~5.2 % of the genome. Of 2,960 SD detected in duck, only 7 exceed 10 kb in length and none is greater than 20kb. On the contrary, large SD are abundant in mammals. It might not be a drawback resulting from the second generation sequencing technique because chicken sequenced by Sanger doesn't have large SD (> 20kb) either. Detailed analyses of SD have shown that a total of 412 genes are located in the predicted duck SDs and 209 of them can be annotated by the Gene Ontology database. Those genes participate in immunity, receptor, and signaling pathways, suggesting that SD plays an important role in the organism's adaptive

evolution. In addition, the genes related to the cytoskeleton are also enriched in the predicted duck SD, which was not reported in other species (Huang *et al, in prep*).

The draft duck genome sequence will allow more detailed comparative analyses to study candidate genes involved in the immune response to avian influenza. Comparison of the turkey, duck, chicken and zebra finch genomes, allowed the identification of 5, 76, 577 and 1,752 lineage-specific gene duplications (LSD). The use of different sequencing platforms may partially explain the lower number of LSD in turkey and duck, but both chicken and zebra finch were sequenced by the Sanger technique, meaning that the difference in LSD could reflect the requirement of gene expansion for adaptation. Within the 76 duck LSD, 14 gene families are found, out of which 3 are significantly expanded in the duck lineage. Those three significantly expanded gene families are: (1) the immunoglobulin superfamily (*IgSF*) which includes mammalian butyrophilin-like (*BTNL*) genes with the exception of *BTNL9*. *BTNL* were suggested to attenuate T cell activation and antagonize the pathologic inflammatory T cell infiltrates (Bas et al. 2011; Nguyen et al. 2006). Inside this superfamily, other *BTN* members and the tripartite motifs (*TRIM*) also exhibit a special domain which are involved in the secretion of lipid droplet (Jeong et al. 2009) or in targeting retroviral capsid proteins(Towers 2007) and binds to the Fc portion of *IgG* (James et al. 2007); (2) an olfactory receptor (*OR*) gene family. This expanded gene family may be a result from the adaptation of aquatic lifestyle for duck comparing to turkey, chicken and zebra finch; (3) a novel gene family that includes only 5 duck epidermal growth factor (*EFG*)-like genes.

Due to the lack of resolution of intermediate map, the duck assembly is much more fragmented than that of chicken or zebra finch. To facilitate comparative genomics, QTL detection and fine mapping in duck, efforts should be devoted to improve the genome assembly and accomplish chromosome assignment.

4.6 Ultrascaffold construction strategy for NGS: duck as an example

The duck sequencing project has produced 78,487 scaffolds; no detailed BAC-based or any other physical map is yet published and only a very low density genetic map is available at the moment (Huang et al. 2006). Thus, with the current data, it is almost impossible to concatenate scaffolds and to assign scaffolds to chromosomes in a correct order. Although the location of scaffolds can be partial inferred through comparative genomics with chicken, due to avian's well-conserved synteny, their orientation and local ordering may sometimes be wrong, as suggested by the few intrachromosomal rearrangement detected in cytogenetic comparative maps (Fillon et al. 2007; Skinner et al. 2009). It is unsuitable to order and orient such a massive numbers of scaffolds by a cytogenetic approach such as FISH alone. Moreover, the resolution of FISH is insufficient.

Thanks to NGS, more and more species will be subjected to genome sequencing, among which most of them may have even less genomic data available than duck. There is no doubt that the genome assembly will be highly fragmented like that of the giant panda and of duck, causing difficulties in constructing chromosome assemblies.

Therefore, we propose here a strategy for improving NGS genome assembly, aiming at building chromosome-wide sequence assemblies. RH mapping can be used to construct NGS chromosome sequence. RH mapping reduces a lot of complexities; moreover, combining NGS makes RH mapping more powerful and high throughput. We use duck as an example, showing that the feasibility of this survey approach. Moreover, by combining comparative genomics information with other sequenced birds, the assignment of duck scaffolds onto chromosomes is achieved and thereafter allowing detection of chromosomal rearrangement among them.

Chapter II.

Construction and Characterization of Duck Whole Genome Radiation Hybrid Panel

1. Introduction

Since the first ones developed in human (Gyapay et al. 1996), whole-genome radiation hybrid (WGRH) panels have been widely produced for mammalian species. The radiation hybrid (RH) maps produced with these panels have usually a higher resolution than the genetic maps produced by recombinant mapping, allowing the ordering of markers otherwise clustered on the genetic map (Gyapay et al. 1996). However, another major advantage of RH over genetic mapping, is that it does not require polymorphism: any STS (Sequence Tagged Site) can be used. This has proved especially useful for mapping genes and EST (Expressed Sequence Tags). The resolution of RH maps can be tailored by adapting the radiation dose used to break the chromosomes in the donor cell. Higher radiation doses will break the chromosomes into smaller fragments and panels of different resolutions can be created depending of the needs: aid to BAC contig construction, high resolution transcript maps of a whole genome, or regional fine mapping of candidate regions for quantative trait loci (QTLs) (Faraut et al. 2009). Radiation hybrids are produced by the fusion of lethally irradiated donor cells of the species of interest with a recipient cell line, usually of rodent origin, which is either thymidine kinase (TK) or hypoxanthine guanine phosphoribosyl transferase (HPRT)-deficient. Fusion products are cultured in selective hypoxanthine-aminopterin-thymidine (HAT) media to eliminate the parental rodent cells and isolate the hybrid clones. The WGRH (Whole Genome Radiation Hybrid) panel obtained consists of hybrid clones that randomly retain a subset of short broken chromosomal fragments from the donor cells. The markers are then scored by a simple PCR analysis for the presence or the absence of DNA from the hybrids, avoiding the necessary development of polymorphism as required for genetic maps (Figure I-10). The probability that two linked markers are included within a single fragment, and therefore their co-retention probability, decreases with the distance between them. This method allows the mapping of a high number of non-polymorphic markers such as expressed sequenced tags (EST) or gene based markers, providing an efficient approach for direct gene

mapping and the production of dense maps of the genome. However, the proportion of the genome from the donor cells -the retention frequency- is critical to the success or failure of a WGRH panel. Hybrid selection was proposed by Jones (1996) as a method to increase the mapping power of radiation hybrid panels (Jones 1996). In that case, several hundred of hybrids are made initially and screened for the proportion of donor cell genome present in the hybrids, assessed by PCR analysis by testing for the presence or the absence of a small number of independent markers to provide independent estimate of donor cell chromosome retention and to assess the genome-wide retention frequency. Then, a selection is made for a subset of 90 hybrids which are positive for the largest proportion of tested loci.

In birds, an attempt to develop chicken radiation hybrids was first published by Kwok and coworkers who tested 4 different radiation doses and two different hamster recipient cell lines but got only a few hybrids in each case (Kwok et al. 1998). This was probably due to a particularly low retention frequency of the chicken genome after the fusions. Indeed, a lower retention of the chicken chromosome fragments leads to a lower number of hybrids bearing the selection gene and thus leads to a lower number of hybrids from which to select after each fusion experiment. To overcome this problem when developing the chicken radiation panel, a large number of fusion experiments was done, to obtain more than 450 chicken radiation hybrids, whose average retention frequency was only 11.3% for the whole genome (Morisson et al. 2002). Due to the particularities of the chicken genome structure, the retention rate for markers located on microchromosomes and macrochromosomes were evaluated separately giving values of 14.8% and 9.5% respectively. Finally, the 90 best clones were selected for the final WGRH panel in which the average retention frequency of the chicken genome is close to 22% (25.7% for the microchromosomes and 20.1% for the macrochromosomes). This WGRH panel has successfully been used to construct chromosome RH maps of the chicken genome and helped in detecting regions misplaced in the sequence (Morisson et al. 2007).

Our experience in chicken highlighted the difficulties in developing such a tool in birds, the main problem being the low retention frequency. In an attempt to produce a WGRH panel in duck and in order to maximize the number of hybrids obtained from each fusion experiment, we decided to optimize our method by comparing the conditions we used in chicken with those recently published by Page and Murphy for mammals (Page and Murphy 2008). It is anticipated that the optimized method described here should be applicable to other birds. Along with duck genome sequencing, in the absence of other long-range intermediate maps in duck, the duck WGRH panel will be the only source to aid in the improvement of current duck genome assembly.

2. Results and discussion

2.1 Comparison of two methods for duck embryonic fibroblast culture

For chicken radiation hybrids, we used normal diploid fibroblasts obtained from female chick embryos and propagated in complete RPMI1640 medium [RPMI1640 (Sigma Chemical Co.) supplemented with 15% fetal calf serum, streptomycin and penicillin] with 5% CO_2 and at 40°C to emulate the natural chicken body temperature. They were cryopreserved in complete RPMI1640 plus 15% glycerin at a concentration of 3 to 6 million cells/mL in liquid nitrogen. However duck primary fibroblasts grew better when cultured in complete DMEM [DMEM plus GlutaMAX^{TM}-I (Gibco/Invitrogen), supplemented with 10% fetal calf serum, 1% streptomycin and penicillin] and cryopreserved in 95% fetal calf serum (Gibco/Invitrogen) and 5% DMSO. These conditions are in accordance with the ones recommended by Page and Murphy.

2.2 Generation of duck radiation hybrids

Pre-irradiation step: Duck fibroblasts were cultured at 40°C with 5% CO_2 in complete RPMI1640 medium and harvested, on the day of fusion by a PBS wash, trypsinized, collected in complete RPMI1640 (supplemented with 10% fetal calf serum, 1% streptomycin and penicillin), counted, spinned, resuspended in incomplete RPMI1640 medium (without serum) and kept on wet ice until the irradiation step. Wg3hCl_2 hamster cells were cultured in complete RPMI1640 medium at 37°C with 5% CO_2 and prepared for fusion the same way in incomplete RPMI1640 medium. Again, these conditions are in accordance with the ones recommended by Page and Murphy except that DMEM medium was used instead of RPMI1640 medium.

Irradiation and fusion steps:

Protocol 1 (used for the chicken panel): the tube containing chicken fibroblasts was kept on ice during the irradiation and the fusion was performed within 30 min. The importance of these conditions was also highlighted by Page and Murphy in Protocol 2. For the duck panel, the irradiated fibroblasts were added to hamster cells at a ratio of 1:1 and the mixed cells were spinned and resuspended in 1 mL of polyethylene glycol 1500 (PEG-1500) in less than 1 min before adding complete RPMI1640 rapidly. The cell suspension was then dispensed in 75 cm^2 flasks with $5X10^5$ cells of each fusion partner per flask. Finally, fused cells are resuspended in complete RPMI and HAT was added to the medium only 24 hours after fusion.

Protocol 2 (described by Page and Murphy): The donor and recipient cells were treated in the same way as in Protocol 1 for the irradiation step. In the fusion process, Page and Murphy recommend to resuspend the mixed cells in PEG-1500 for a total of 2 min and to add 10 mL of unsupplemented DMEM at a rate of 1 mL/min before centrifuging the cell suspension at 67g for 5 min. Then the cells are resuspended in unsupplemented DMEM and placed at 37°C with gentle mixings every 20 min before being spinned again at 185g for 5

hybrid	Retention	Total # Chr	# micro
h158	16,1%	29	10
		29	9
		29	10
		29	10
		29	9
		29	9
		29	8
		29	8
		29	7
		29	6
h165	9,7%	23	3
		22	0
		22	2
		polyploidy	-
		22	0
		23	0
		28	5
		25	3
		22	1
		24	3
		26	4
h219	25,8%	25	5
		32	12
		26	7
		27	6
		29	6
		25	5
		28	7
		25	4
		26	7
		26	4
h279	32,3%	31	7
		29	7
		31	7
		29	9
		31	8
		28	3
		34	12
		35	9
		34	9
		36	10

Table II-1: chromosome counting results for the four investigated hybrids.
For each hybrid, chromosome counting were carried out on 10 cells. The retention rate was given for each hybrid. The total chromosome number was given in the 3rd column, microchromosome number was given in the 4th column. Compare with the recipient cell karyotype, we observed that the total chromosome number is increased and we detected additional microchromosomes. This study gave an evidence that hybrid cell lines were a mixture of cell population. Furthermore, the total chromosome number had tendency of being positively related with retention rate.

min. Finally, fused cells are resuspended in complete DMEM and cultured directly in presence of HAT (5 x 10^5 cells per culture dish) after fusion.

2.3 Comparative results

Altogether, the main differences between our method and that of Page and Murphy concern the fusion step. The cell partners are submitted to 2 centrifugations according to Page and Murphy's method while we do not spin the cells after addition of PEG-1500. Moreover, in our method, the HAT is added 24 hours later than in Page and Murphy's (see Table II-1). In order to compare the two methods (Page and Murphy's one versus ours) and 2 temperature conditions to cultivate the hybrids (37°C versus 40°C), we carried 2 fusion experiments and for each tested 4 fusion conditions (combination of two protocols and two culture temperatures), each using 17 million of both cell partners in each case. No clone was found at 40°C either with Page and Murphy's method or with ours. Fifteen clones were observed and cultured using our method and 3 clones were observed and cultured using the one of Page and Murphy. Taking these results into account, we set up a large scale fusion according to the method described below.

2.4 The optimized method

2.4.1 Primary fibroblast culture and cryopreservation

Twelve-day duck embryos were chosen as donors of primary fibroblasts. The eggs were washed with 70% ethanol and the embryos were carefully picked with tweezers and placed in culture dishes. They were washed twice with a 0.05% trypsin solution (comprising 8g NaCl, 0.4g KCl, 1g glucose, 0.58g $NaHCO_3$, 0.2g EDTA and 0.5g Trysin per 1 liter solution). The heads of the embryos were removed and the embryos were eviscerated before being dilacerated. The trypsin digestion of the tissues was carried out in 5 mL of 0.3% trypsin solution (increasing trypsin to 3g without EDTA compared to 0.05% trypsin solution) at 37°C

for 5 min under gentle stirring. The supernatant was poured into a 50 mL tube containing 10 mL complete DMEM to stop the trysinization. This step was done three times and the cell suspension was spinned before being resuspended in complete DMEM and counted. They were used to set up the primary fibroblast culture while the rest of the tissues was kept for DNA extraction in order to determine the sex of the embryos by PCR amplification according to Batellier et al(Batellier et al. 2004). Indeed, the fibroblasts will have to be female which is the heterogametic sex in birds. Fibroblasts were cultured in complete DMEM [DMEM Glutamax (Gibco Co.) supplemented with 10% fetal calf serum (Gibco Co.), 1% penicillin and streptomycin] at 40°C with 5% CO_2 in 25-cm^2 flasks on the basis of 6 x 10^6 cells/flask. At the stage of confluence, the cells were harvested, centrifuged, resuspended in 95% fetal calf serum (Gibco Co.) and 5% DMSO, at a concentration of 6 to 12 million cells/mL and cryopreserved in liquid nitrogen.

2.4.2 Generation of radiation hybrids

Pre-irradiation step: Two 75 cm^2 flasks were seeded with 6 million female duck fibroblasts in complete DMEM and cultured in 5% CO_2 at 40°C until the cell monolayer approached confluency. The day before the fusion, the cells were harvested, divided in four 75 cm^2 flasks and cultured overnight. On the day of fusion, the medium was aspirated from the flasks and the monolayer was washed twice with 10 mL of HBSS. Trypsinization in each flask was carried out using 2 mL of 0.05% trypsin solution, and the fibroblasts were collected in 5 mL of complete DMEM. All the fibroblasts from the 4 flasks were pooled in the same tube before centrifugation for 10 min at 900 rpm. The fibroblasts pellet was resuspended in 15 mL of complete DMEM and this centrifugation step was repeated to remove all trypsin from the solution. The fibroblasts were then resuspended in incomplete DMEM (without serum) and counted. One tube containing 27 million fibroblasts was kept on wet ice until the radiation step. Wg3hCl_2 hamster cells were cultured the same way in complete RPMI1640 medium at

Parameters	Protocol 1	Protocol 2
Donner cell treatment (pre-irradiation)	Incomplete DMEM	Incomplete DMEM
recipient cell treatment (pre-irradiation)	Incomplete DMEM	Incomplete DMEM
PEG treatment	1 min	2 min
Resuspension media	RPMI1640	DMEM
Centrifugation after fusion	Not needed	Yes, 67g for 5 min
Incubation of fusion partners	Not needed	Incubate at 37°C for 1 h with one gentle mixing every 20 min, and then cells were centrifuged before division
Addition of HAT	24 hrs after fusion	Immediately
Media for hybrid culture	Complete RPMI1640 with HAT	Complete DMEM with HAT

Table II-2: **Comparison of two protocols used for generation of radiation hybrid.**
Protocol 1 was adapted for chicken RH panel construction whereas protocol 2 was described by Page and Murphy (2008). Incomplete DMEM only contains 1% streptomycin and penicillin (without serum). Complete DMEM/RPMI medium are supplemented with 1% streptomycin and penicillin and 10% fetal calf serum.

37°C with 5% CO_2. Twenty-seven million cells were prepared in complete RPMI1640 medium

Irradiation and fusion steps: The tube containing 27 million duck fibroblasts was kept on ice while being irradiated at 6,000 rads. The fusion step was carried out within the 30 min after the irradiation step. In the fusion process, twenty five million irradiated fibroblasts were mixed to 25 million hamster cells and spinned at 900 rpm for 10 min. The pellet was resuspended in 1 mL PEG 1500 (Roche Applied Science) in less than 1 min (checked with a timer) by gently pipetting up and down with a 1-mL pipette. Twenty-four mL of incomplete DMEM (without serum) was rapidly added. The cell suspension was then dispensed in fifty 75 cm^2 flasks (0.5 million of each fusion partners per flask) in RPMI1640 and placed at 37°C with 5% CO_2. Two million hamster cells were seeded in one 75 cm^2 flask containing 20 mL of complete RPMI (negative control N°1) and 2 million irradiated fibroblasts were seeded in one 75 cm^2 flask containing 20 mL of complete DMEM (negative control N°2).

Post-fusion step: HAT (HAT Media Supplement (50×) Hybri-Max® from Sigma-Aldrich), was added 24 hours after the fusion in all the flasks except in the negative control N°2. The medium was changed (complete RPMI1640 plus HAT) in all the flasks 4 days post-fusion to discard the non-fused cells and once a week afterwards. The flasks were examined for the presence of hybrid clones everyday between 7 and 20 days after fusion.

2.4.3 Clone picking, short term cultures and DNA extraction

No colonies were observed either in negative control N°1 or N°2, indicating that only the fusion products were viable under the combination of irradiation and selective medium conditions used. Seven to 12 days after the fusion, clones appeared, sometime several of them in the same flask. After the clones had grown enough to occupy the whole field of the microscope (objective 10X), they were individually picked using bent Pasteur pipettes and transferred to individual 25cm^2 flasks. Each clone was cultured until the stage of confluence

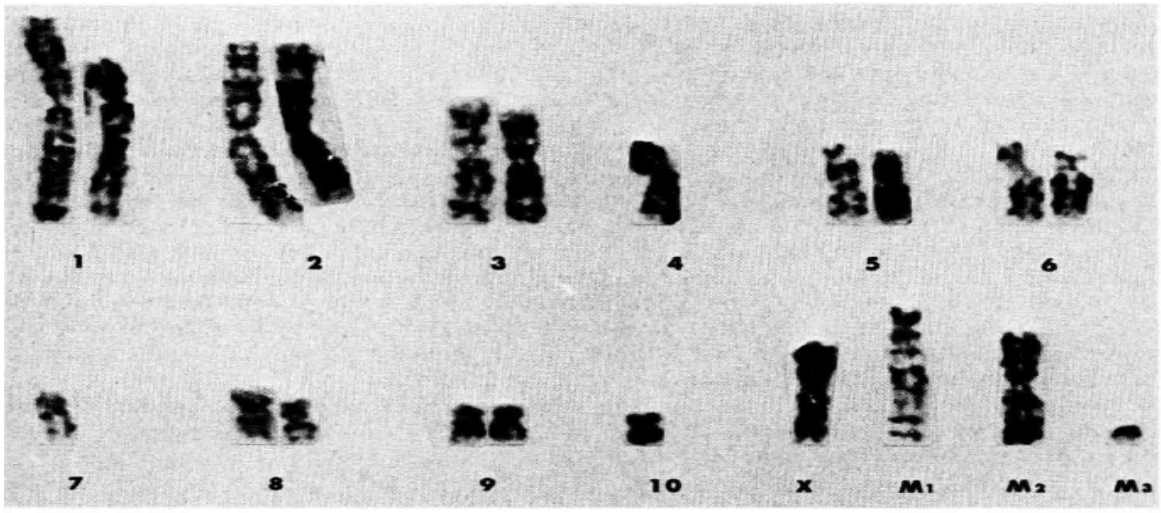


Figure II-1: karyotype of Wg3hCl_2 cell lines. The cell line derives from Chinese Hamster Lung cells (DON) and was characeterised by Echard *et al*. (1984). Due to karyotype instability, the chromosome number can vary between 20-24 with a median value of 21.

and transferred in two 75cm^2 flasks. Clones appeared at various times throughout the 7-20 days post fusion and grew at different speed. Care given to each clone was adapted to its own behaviour, the medium being changed at least once a week and trysinization steps were added for clones growing in lumps, in order to reseed the whole flask. When fully grown, cells from both flasks were harvested. Five million were kept for DNA extraction while the rest was frozen in 95% fetal calf serum (Gibco Co.) and 5% DMSO, at a concentration of 3 to 6 million cells/mL and cryopreserved in liquid nitrogen**.** DNA extraction was carried out using the QIAGEN DNeasy® blood and tissue kit (http://www1.qiagen.com/Products/).

2.5 Cytogenetic investigations on four hybrids

We randomly chose 4 duck RH hybrids (h158, h165, h219 and h279) to establish the chromosome number and visualize the location of duck genome fragments by FISH (Fluorescent *In Situ* Hybridisation).

Chromosome number

We counted 10 metaphases for each hybrid. Results are shown in Table II-2. Karyotype of Wg3hCl$_2$ hamster cells had been analyzed by (Echard G. 1984) as shown in Figure II-1**.** This cell line has a chromosome number ranging from 20 ~ 24 with a median value of 21 (Echard G. 1984) and exhibits a very small chromosome M3 (Figure II-1). For all the four hybrids, the number of chromosomes is increased compare to the Wg3hCl2, showing additional microchromosomes. We observed a strong variability of the number of chromosomes between and within the hybrids. Although h158 showed constant chromosome number in all 10 cells checked, microchromosome number differed in which some contained 6 microchromosomes and 3 or 4 very small fragments and some contained 5 microchrosomes. The remaining three hybrids had a variable number of chromosomes according to metaphases. The retention rate seemed somehow positively related with the total chromosome number: the

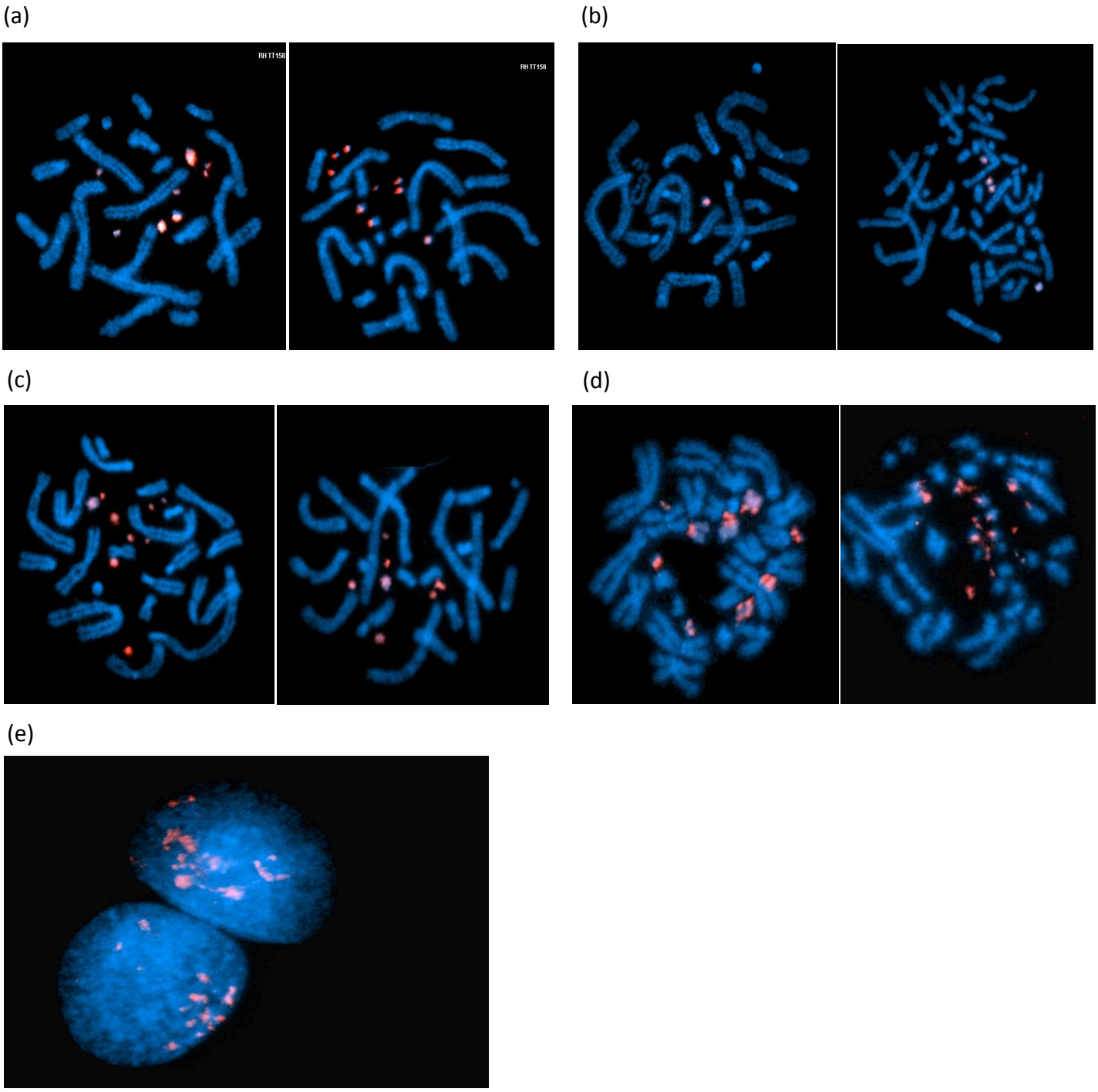


Figure II-2 : Cytogenetic study of 4 duck hybrids. Duck genomic DNA (red) is hybridised to hamster chromosomes stained by DAPI (blue). (a) two metaphases from hybrid h158, with 8 (left) and 9 (right) additional microchromosomes containing duck fragments. Moreover, the size of the microchromosomes is different between these two cells. This suggest that the hybrid cells are not monoclonal. (b) two examples from hybrid h165, with only 1 (left) and 4 (right) additional chromosomes. The cell on the right is tetraploid. (c) two examples of hybrid h219 and (d) two examples of h279. Cells contain different number of additional microchromosmes. (e) a 2D-view of the duck fragments in hybrid interphase nuclei.

higher the retention rate, the higher the chromosome number. This suggests that the fragmented duck DNA could integrate the recipient cells as independent neo-microchromosomes. The different chromosome numbers suggests that there is a strong instability of the karyotype in the hybrid cells.

FISH of duck genomic DNA

To further visualize the integration of the fragmented duck DNA in the hybrid cells, labeled duck genomic DNA was used as a probe and hybridized *in situ* on metaphase chromosome spreads of hybrid cells (Details see in supplementary method). FISH experiments showed that the duck chromosome fragments formed one or more additional microchromosomes and in some cases a few small fluorescent signals are seen on hamster chromosomes, suggesting that the duck chromosome fragments preferentially form additional microchromosomes and only occasionally insert themselves in the hamster chromosomes (Figure II-2). Meanwhile, within the same hybrids, we observed that the additional microchromosomes containing duck fragments could vary in size and number, which was consistent with the observation above in counting chromosome numbers. This study provided further evidence that a hybrid cell line was not monoclonal but a mixture of a cell population (V.Fillon, unpublished data).

2.6 Discussion

The interspecific hybrids are obtained by fusing cells that grow at different temperatures: 37°C for the hamster cells and 40°C for the duck cells. The question is thus whether to use a temperature that favors the donor or the recipient cells. Favoring the donor cells could perhaps help for chromosome fragment retention, whereas favoring the recipient should ensure optimal growth, as their genome is complete, unlike the donor cell's genome. Previous studies were made on somatic hybrids (Cassingena et al. 1971; Grzeschik et al. 1972; Kao 1973; Migeon and Miller 1968; Minna and Coon 1974; Minna et al. 1974; Westerveld et

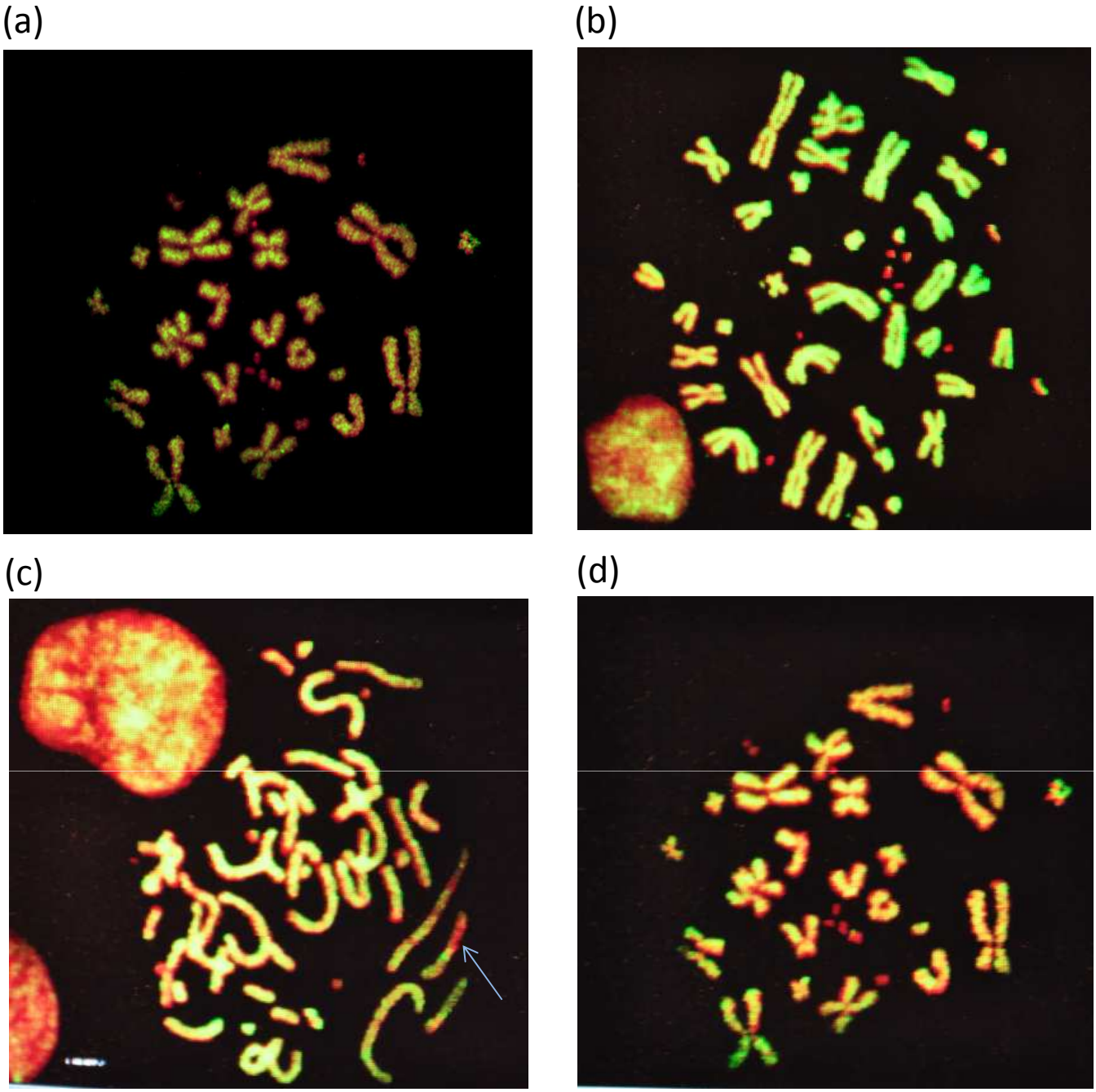


Figure II-3 : Cytogenetic study of Chicken hybrids using primed in situ labelling (PRINS) of the hamster genome. Green/yellow PRINS signals on red stained chromosomes are from hamster and non-labelled red chromosomes are from chicken. Most chicken fragments are in the form of additional microchromosomes, with only very few insertion in the hamster genome, indicated by the arrow in (c). (V.Fillon personnel communication)

al. 1971), showing that genome dominance was observed. Herein, we suggest the recipient cell should dominate the cell cycle, at least from the fact that donor cells are only partially retained and thus they don't have a whole set of genes for cell growth and propagation. Work on the zebrafish RH panel also reflected that culture conditions should be those of the recipient cell (Kwok et al. 1998). The difficulty for obtaining more chicken hybrids is probably due to the fact that the chicken hybrids were cultured at 39°C instead of 37°C.

Apart from the culture temperature, the fusion process is also a critical factor influencing fusion efficiency. The protocol described by Page and Murphy could be very stringent for fragile irradiated cells; especially the long treatment with PEG followed by centrifugation in the post irradiation step may be fatal for the fragile hybrids.

The painting experiment of duck genomic DNA onto the hybrids showed the presence of donor cell chromosome fragments as additional chromosomes independent from the hamster chromosome. This is similar to what had been observed for chicken hybrids (Kwok et al. 1998) and also in agreement with the characterization of chicken hybrids where the chicken DNA fragments were formed as independent microchromosomes (V.Fillon personnal communication, Figure II-3). The painting on chicken hybrids had been done by PRINS (Primed in situ labeling) of hamster fragment, and showed only occasional cases (Figure II-3.C) of chicken fragments inserted into hamster chromosome and mostly independent additional microchromosomes. These cytogenetic studies have shown that duck and chicken hybrids behave in a similar pattern. However, these observations are slightly different from those in swine, for which larger additional chromomes were observed (Yerle et al. 1998) (Figure II-4) and in zebrafish for which insertions into the recipient genome were not so rare (Kwok et al. 1998). As a conclusion, what can be deduced from these cytogenetic studies, is that the donor cell fragments frequently form additional microchromosome(s) or chromosome(s) which are retained in the hybrids, at least for chicken and duck hybrids. Combining the observation that an additional chromosome could consist of fragments coming

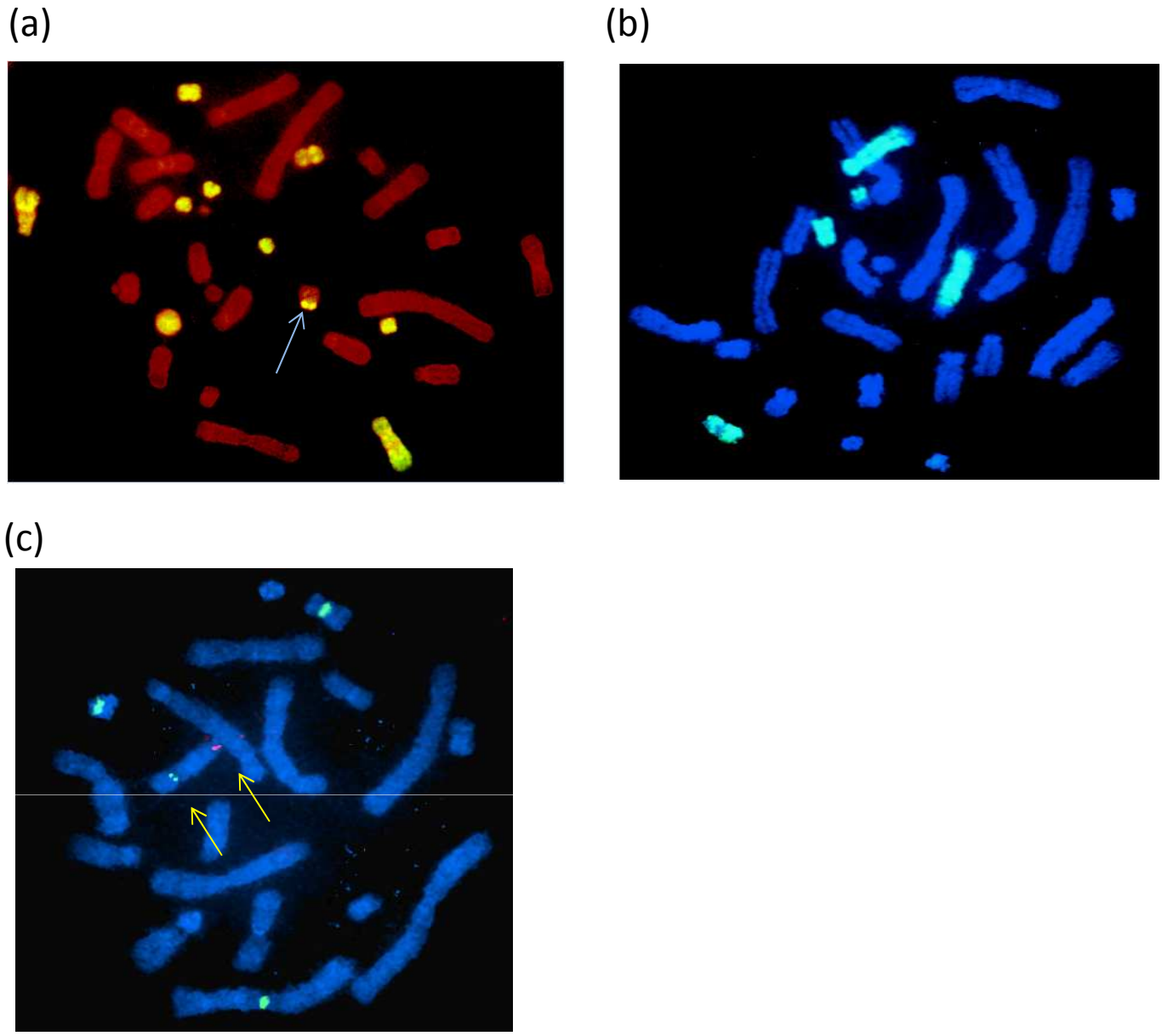


Figure II-4 : (a) and (b) characterization of swine hybrids (6000 rads panel) using primed in situ labelling (PRINS) of the swine genome. (a): Green/yellow signals on red chromosomes are from swine fragments labelled by PRINS on red (a) or blue (b) hamster chromosomes. Results suggest that swine fragments could form additional chromosomes, or insert themselves in hamster chromosome (arrow in (a)). (c) the centromeric sequences of swine chromosomes were labeled as probes. Probes derived from metacentric and acrocentric chromosomes were labeled by different colors. Two signals in (c) suggest this chromosome contains fragments from at least from two different chromosmes. (M.Yerle personnal communication)

from different donor cell chromosomes (Kwok et al. 1998; Yerle et al. 1998), we postulate that the additional chromosomes are comprised of the joining of many small radiation-induced duck chromosome fragments with randomly selected duck centromeres, which are required for proper segregation during mitosis. This is reflected by the higher retention of centromeric regions of donor chromosomes in the hybrids (Figure II-5). However, the observation from swine hybrids (Figure II-4.c) and cytogenetic study by Yerle *et al*.(1998), that multi-chromosomal originated additional chromosome has more than one centromere, raises an interesting question that what mechanism would be involved to inactivate extra centromere(s) if there is more than one, otherwise chromosome would not be stable.

Some additional chromosomes observed in swine hybrids could be much larger than that of in birds as represented in Figure II-2, 3, 4, and insertion events were more common in zebrafish hybrids (Kwok et al. 1998). The mechanism is not clear yet, but we speculate that the repetitive sequences might be involved as half of swine and zebrafish genomes are repetitive sequences and containing much more DNA transposons than birds (Lam et al. 1996; Wiedmann et al. 2006).

3. Conclusion

This feasibility study compared four fusion conditions and established an optimized protocol to generate radiation hybrids in birds. We carried out two fusion experiments to test all 4 conditions from which the optimized condition provided highest fusion efficiency. Additional fusion experiments with the best condition, using the same protocol as for chicken radiation hybrids described by (Morisson et al. 2002), but with a culture temperature of 37 ^{o}C will be needed to generate enough hybrids for the final panel.

Centromere effect

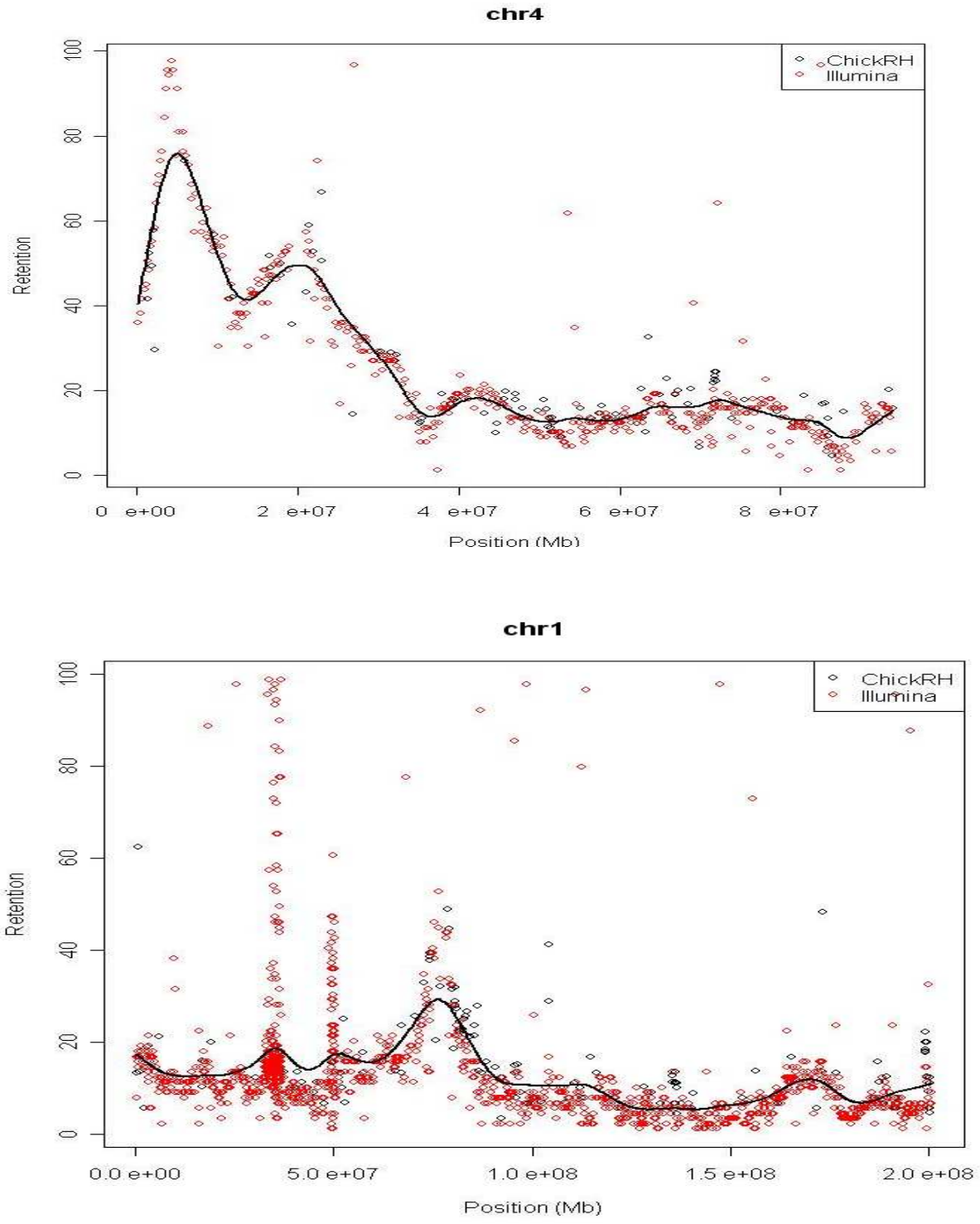


Figure II-5: Retention values in the chicken ChichRH6 panel. Top: along GGA4. The *HPRT* gene used for donor chromosome fragment selection is located near the first peak (100% retention). The second peak is close to the centromere. Bottom: along GGA1. (A.Vignal personnel communication)

4. Supplementary Method

Fluorescent in situ hybridisation (FISH)

FISH was carried out on metaphase spreads obtained from cell cultures of four RH-hybrids, arrested with 0.05 µg/ml colcemid (Sigma) and fixed by standard procedures. Duck genomic DNA was extracted from cryopreserved cells (5 million) deriving from the same individual (TT) that was used for the panel construction, with the Qiagen Dneasy blood and tissue kit (spin-column protocol).

The FISH protocol is derived from Yerle et al, 1992. Two-colour FISH was performed by labelling 100 ng of TT genomic DNA with alexa fluorochromes (ChromaTide® Alexa Fluor® 568-5-dUTP, Molecular Probes) by random priming using the Bioprim Kit (Invitrogen). The probe was purified using spin column G50 Illustra (Amersham Biosciences), ethanol precipitated, and resuspended in 50% formamide hybridization buffer. The probe was denaturated and hybridised to RH hybrids metaphase slides for 17 hours at 37°C in the Hybridizer (Dako). Chromosomes were counterstained with DAPI in antifade solution (Vectashield with DAP, Vector). The hybridised metaphases were screened with a Zeiss fluorescence microscope and a minimum of twenty spreads was analysed for each experiment. Spot-bearing metaphases were captured and analysed with a cooled CCD camera using Cytovision software (Leica-Microsystem).

ChapterIII.

Testing the Duck RH panel with Different Genotyping Techniques

Introduction

After our tests on culture and fusion conditions, we obtained an optimized protocol to construct a radiation hybrid (RH) panel in duck. However, during our test studies, we couldn't obtain enough hybrids for a whole panel. Therefore, two more fusion experiments were carried out and a total of 225 hybrid clones were harvested from four fusion experiments in total. From these, a selection of a set of the 90 best hybrids, with the highest retention values for duck chromosomes, is mandatory for increasing the RH mapping power (Jones 1996).

To assist the duck genome sequence assembly, thousands of markers will have to be genotyped on the panel, meaning that a large quantity of DNA from the hybrids is needed. For this, large scale culture of the hybrid clones is the usual approach. However, this is a time-consuming step and another major problem is that donor chromosome fragments are lost from the hybrids in the process. To avoid this, we first tried using whole genome amplification (WGA) to amplify all the hybrids in the panel and assessed the retention variation before and after WGA with a same set of microsatellite markers chosen all over the genome on the basis of existing genetic maps. Then, to investigate the power of the panel for building maps, we developed markers from the duck assembled scaffolds and genotyped both on WGA panel and non-WGA panel.

Traditional RH mapping involves genotyping defined markers by PCR followed by migration on agarose gel. Although it proved effective for building maps in many species, this method is time-consuming for genotyping high numbers of markers. Therefore, we tested the possibility of using the Fluidigm real time quantitative PCR (qPCR) (Spurgeon et al. 2008) who allows the genotyping of 96 markers at a time and a drastic reduction in reaction volumes. These volumes being as low as 7nL, thousands of markers can be genotyped even without large-scale culture of the clones. Therefore, we tested the Fluidigm real time qPCR on both the WGA panel and on the non WGA panel.

The different genotyping methods were assessed with a set of 39 markers and we selected the method having the best performance to evaluate the quality of the duck genome assembly. This was the first assessment study of a NGS assembly by RH genotyping so far. Using the Fluidigm real time qPCR techniques, we genotyped duck EST markers showing no BLAST hit to the chicken genome, as these could correspond to microchromosomes or other genomic sequence absent from the assemby. The result show that RH mapping by Fluidigm qPCR are more powerful than traditional PCR and agarose gel electrophoresis genotyping. This might thus help assembling the sequence of the 10 smallest michrochromosomes, still causing problems in chicken and quite certainly also in duck.

Article

Insert article" **A duck RH panel and its potential for assisting NGS genome assembly"**

Accepted by BMC Genomics

A duck RH panel and its potential for assisting NGS genome assembly

Man Rao[1], Mireille Morisson[1], Thomas Faraut[1], Suzanne Bardes[1], Katia Fève[1], Emmanuelle Labarthe[1], Valérie Fillon[1], Yinhua Huang[2], Ning Li[2], Alain Vignal[1*]

1-UMR INRA/ENVT Laboratoire de Génétique Cellulaire, INRA, Castanet-Tolosan, 31326, France

2-State key laboratory for agro-biotechnology, China Agricultural University, Beijing 100193, People's Republic of China

* Corresponding author

Email addresses:

MR: rao.man@toulouse.inra.fr

MM: mireille.morisson@toulouse.inra.fr

TF: thomas.faraut@toulouse.inra.fr

SB: suzanne.barde@toulouse.inra.fr

KF: katia.feve@toulouse.inra.fr

EL: emmanuelle.labarthe@toulouse.inra.fr

VF: valerie.fillon@toulouse.inra.fr

YH: yinhuahuang@126.com

NL: ninglcau@cau.edu.cn

AV: alain.vignal@toulouse.inra.fr

Keywords: RH mapping – NGS – Sequencing – Duck – Scaffold - Assembly

ABSTRACT

Background: Owing to the low cost of the high throughput Next Generation Sequencing (NGS) technology, more and more species have been and will be sequenced. However, *de novo* assemblies of large eukaryotic genomes thus produced are composed of a large number of contigs and scaffolds of medium to small size, having no chromosomal assignment. Radiation hybrid (RH) mapping is a powerful tool for building whole genome maps and has been used for several animal species, to help assign sequence scaffolds to chromosomes and determining their order.

Results: We report here a duck whole genome RH panel obtained by fusing female duck embryonic fibroblasts irradiated at a dose of 6,000 rads, with HPRT-deficient $Wg3hCl_2$ hamster cells. The ninety best hybrids, having an average retention of 23.6% of the duck genome, were selected for the final panel. To allow the genotyping of large numbers of markers, as required for whole genome mapping, without having to cultivate the hybrid clones on a large scale, three different methods involving Whole Genome Amplification (WGA) and/or scaling down PCR volumes by using the Fluidigm BioMark™ Integrated Fluidic Circuits (IFC) Dynamic Array™ for genotyping were tested. RH maps of APL12 and APL22 were built, allowing the detection of intrachromosomal rearrangements when compared to chicken. Finally, the panel proved useful for checking the assembly of sequence scaffolds and for mapping EST located on one of the smallest microchromosomes.

Conclusion: The Fluidigm BioMark™ Integrated Fluidic Circuits (IFC) Dynamic Array™ genotyping by quantitative *PCR* provides a rapid and cost-effective method for building RH linkage groups. Although the vast majority of genotyped markers exhibited a picture coherent with their associated scaffolds, a few of them were discordant, pinpointing potential assembly errors. Comparative mapping with chicken chromosomes GGA21 and GGA11 allowed the detection of the first chromosome rearrangements on microchromosomes between duck and

chicken. As in chicken, the smallest duck microchromosomes appear missing in the assembly and more EST data will be needed for mapping them. Altogether, this underlines the added value of RH mapping to improve genome assemblies.

BACKGROUND

The development and commercialization of next-generation massively parallel DNA sequencing approaches, by dramatically decreasing the cost of sequencing, have revolutionized genomic research. The main innovation of NGS, as compared to Sanger sequencing, is the parallelisation of the process, allowing between a few thousands and up to millions of sequencing reactions to be processed simultaneously. The three main NGS technologies available on the market use different approaches for library construction, template immobilisation and sequencing reaction, but the basic principles remain the same. NGS approaches also have some drawbacks compared with Sanger sequencing: (1) sequence reads produced currently by NGS (100 bp for Illumina, 500 bp for 454) are shorter than Sanger sequencing reads (1000 bp) and have a higher error rate, making the sequence assembly more problematic; (2) the pairing of reads in Sanger sequencing is limited by the size of the DNA fragments that can be inserted in cloning vectors, ranging from 1-2 kb or less up to 100-200 kb (plasmids, fosmids, BAC), whereas pairing of reads is limited to 40 kb with NGS, limiting the average assembled scaffold length and leading to more difficulty in segmental duplication and copy number variation detection; (3) as a consequence, a higher sequencing depth is required for assembly and a very high number of small scaffolds are produced. Nevertheless, the sequencing and *de novo* assembly of a Chinese individual [1] proved the feasibility of sequencing and assembling whole genomes by NGS. Many species have been sequenced and/or resequenced by NGS, such as the giant panda *Ailuropoda melanoleura* [2], the silk worm *Bombyx mori* [3], the cucumber *Cucumis sativus* [4], the chicken *Gallus gallus domesticus* [5, 6], the turkey *Meleagris gallopavo* [7] and the Mallard duck *Anas platyrhybchos domesticus* (Huang *et al*, in prep).

The Pekin duck (*Anas platyrhynchos*, APL) is an obvious target for detailed genomic studies due to its agricultural importance [8-10] as well as for its role as a natural reservoir of

all influenza A viruses. It can usually carry the infection with no sign of disease and thus propagates the virus to other bird species and potentially to mammals such as pigs or humans [11-15]. The duck genome presents most of the characteristics encountered in birds, which are: (i) a more compact genome, one third the size of a mammals, (ii) a large number of chromosomes (2n = 80), (iii) the presence of macrochromosomes and microchromosomes, the latter being as small as a few Mb [16] and (iv) the females are the heterogametic sex (ZW) and males the homogametic one (ZZ). Due to its importance both in the economic and scientific fields, the sequencing of the Pekin duck genome was initiated in 2008 using the same strategy recently published for the giant panda [2] at the Beijing Genomics Institute (BGI). The sequence reads provided a depth of 65X and a total of 78,487 scaffolds were assembled in which N50 scaffold was 1.2Mb and the largest was 5.9 Mb in length (Huang *et al*, in prep). However, owing to the lack of a clone-based physical map and other supplementary mapping data, apart from a first generation genetic map composed of 155 microsatellite markers, 115 of which located in only 19 linkage groups spanning 1353.3cM [17], it is possible to assign only very few assembled scaffolds to chromosomes.

Several studies have shown that birds seem to have a slower rate of chromosome rearrangements than mammals, with only very little inter-chromosomal rearrangements [18-23]. Between chicken and zebra finch, whole genome comparison revealed 114 tentative intrachromosomal rearrangements (56 inversions and 58 translocations) in which some were confirmed by FISH (Fluorescent *In Situ* Hybridization) experiments [24, 25]. Recently, Zhang *et al* (2011) provided confirmed evidence for 20-27 major rearrangements between turkey and chicken, almost all of which are inversions [26]. The mean reported phylogenetic distance between chicken and turkey is 47 million years, whereas it is 81 between chicken and duck [27], so the number or rearrangements reported between chicken and turkey provide the minimum level of difference expected between chicken and duck. To date, only one

interchromosomal difference has been reported between the chicken and the duck karyotypes, with APL4 and APL10 corresponding to GGA4q and GGA4p respectively [21]. This interchromosomal rearrangement explains the difference in diploid chromosomal number between the two species, which is 2n = 78 in chicken and 2n = 80 in duck and therefore the nomenclature for numbering the duck chromosomes follows mainly that of chicken. Macrochromosomes APL1 to APL9 correspond to GGA1 to GGA9, then APL10 corresponds to GGA4p and finally, the rest of the karyotype is offset by one, with GGA10 corresponding to APL11 and so on. Cross-species fluorescent in-situ hybridization (FISH) studies using chicken BAC clones on duck metaphase spreads showed only a few large scale intrachromosomal rearrangements concerning the largest chromosomes [23, 28]. All this demonstrates a high karyotype stability despite 80 million years of divergence between the two species [29, 30].

As a first attempt to order the duck sequence scaffolds, we aligned the 7,205 ones larger than 1kb to the current chicken assembly using the Narcisse database [31, 32] and successfully positioned 1,787 of them. This still leaves a large number of scaffolds to assign and also means that the ordering of the duck scaffolds and genes we obtained will follow the chicken genome and will be wrong whenever large- or small-scale rearrangements will have happened between the two species.

To assemble the scaffolds in an order corresponding to the real duck chromosomes, several approaches can be used. High density SNP genetic maps allow high precision in mapping. However, the SNP markers need first to be discovered by a sequencing approach, such as published by Kraus *et al*. (2011) [33] and must be informative in a reference population to be used for mapping. However, despite several thousand SNP discovered to date [33], only a small subset of 384 were genotyped [34], mainly due to the high cost that would have been required for additional markers. Finally, out of these, only a small subset

was informative in mapping populations (R. Kraus, R. Crooijmans, personal communication), which will allow only for low resolution maps and poor marker ordering. Furthermore, for high precision mapping, very large populations counting several hundred individuals are required, yet again increasing cost and labour. Further sequencing for SNP detection and a consortium for generating a SNP chip would help improve genetic maps in duck and may happen in the future. Physical maps can be based on the mapping of BAC clones by FISH for chromosome assignment and large-scale ordering. The BAC clones from large libraries can be used for contig construction by fingerprinting or high throughput hybridization. End-sequencing of the clones allow linking sequence scaffolds together. BAC contig maps are thus usually a backbone to the sequence assembly. A BAC library has been made for Duck [35], but to the best of our knowledge, there are no plans yet to build physical maps. In this context, radiation hybrid mapping can be an excellent complementary mapping approach, as it does not require complex marker development and large-scale genotyping. Any STS can be placed on the map by simple PCR on as little as 90 hybrids. Thus with a minimal effort, maps with a resolution intermediate between the genetic and the BAC contig maps can be constructed to propose a correct chromosomal assignment and ordering of scaffolds. We report here the production of a duck whole genome radiation hybrid panel and demonstrate its utility to verify the quality of sequence scaffolds and for assigning and positioning scaffolds onto chromosomes. Large-scale culture of radiation hybrid clones is a time-consuming process and moreover causes the loss of donor fragments in the hybrids. To avoid the necessity of cultivating the radiation hybrid clones at a large scale, we tested three approaches. One involves whole genome amplification (WGA) and conventional genotyping by PCR and gel electrophoresis and the other two use minute amounts of DNA and Fluidigm BioMark™ IFC Dynamic Array™ genotyping by quantitative *PCR*. The advantage of the Fluidigm approaches is low cost combined with simple and rapid high-scale genotyping.

RESULTS

Generation and characterisation of a duck radiation hybrid panel

Duck radiation hybrids were obtained by fusing female duck embryonic fibroblasts irradiated at a dose of 6,000 rads, with HPRT-deficient hamster cells from the $Wg3hCl_2$ cell line. Five fusion experiments were carried out to produce 225 duck radiation hybrids, suggesting that one hybrid clone was recovered per 289,000 duck fibroblasts, corresponding to an average fusion efficiency of 3.46×10^{-6} clone per duck fibroblast. Retention frequencies in the hybrids were estimated by using a set of 31 microsatellite markers distributed along the duck genome, whose positions were estimated on the basis of a low resolution genetic map (Marie-Etancelin *et al.*, in prep). Genotyping was performed by conventional PCR followed by agarose gel electrophoresis. As the ancestral chromosomes 4 and 10, fused in chicken to give GGA4q and GGA4p respectively, remain separated in duck as APL4 and APL10 [23, 28], care was taken to choose one marker located on APL10 and 3 others on APL4. As microchromosomes and the regions close to centromeres were reported to be better retained in chicken radiation hybrids [36-38], we decided to focus more on macrochromosomes and a higher proportion of markers from macrochromosomes. Altogether, 20 markers from macrochromosomes 1 to 7 and chromosome Z were selected and the rest (11 markers) were from identified microchromosomes. By using the genetic maps and comparative mapping with chicken, we avoided the clustering of markers.

As a result from genotyping, we estimated the average retention frequency of duck genome fragments in the 225 hybrids to be 15.3% for the whole genome, with unequal values for macrochromosomes (10.2%) and microchromosomes (21.8%). Previous estimation showed that a panel of 100 hybrids with marker retention frequencies between 20 and 50% are sufficient to build maps of chromosomes at a reasonable resolution [39]. Almost 50% of

our duck-hamster hybrids have an average retention frequency over 15%, being thus potential candidates for the final panel. Finally, the 90 hybrids selected for the definitive panel were chosen with the highest possible marker retentions for macrochromosomes while maintaining good values for microchromosomes. Final retention frequency values are 23.6% genome-wide, with specific values of 20.2% for the macrochromosomes and 28.1% for the microchromosomes.

Testing three different RH strategies for mapping macrochromosomes and medium size microchromosomes

Several thousand markers are needed to build genome-wide maps, requiring large amounts of DNA, usually prepared by large-scale culture of the radiation hybrid clones. However, this is a time-consuming task and moreover, donor chromosome fragments are lost during the culture process. To avoid this, we tested three alternative methods allowing minute amounts of DNA from the hybrids to be used. These were based either on whole genome amplification (WGA) by Multiple Displacement Amplification (MDA) of the DNA from the hybrid clones and/or on scaling down the PCR to 7 nl, allowing the DNA requirements to be as little as 70 pg, by using the Fluidigm BioMark™ IFC Dynamic Array™ genotyping by quantitative PCR (FLDM) [40]. The three conditions were thus: (i) WGA-PCR, in which the DNA from the hybrid clones was amplified by WGA and the genotyping performed by conventional PCR followed by agarose gel electrophoresis; (ii) WGA-FLDMqPCR, in which the WGA-amplified DNA was used for genotyping by quantitative real-time PCR in 7 nl reaction volumes and (iii) Pre-ampFLDMqPCR, in which the DNA from the clones was used directly without WGA, which was replaced by a more specific pre-amplification step using the 96 primer pairs for the 96 loci studied in one Fluidigm BioMark™ run (see methods).

Whole genome amplifications with MDA were performed for all the 90 selected radiation hybrids. Each sample was amplified in three replicates which were pooled together

to avoid representation bias in the final working panel. We obtained a 1000 X amplification efficiency, with more than 150 µg of WGA-DNA obtained for each hybrid, from 150ng of starting DNA. As the smallest microchromosomes have always proved difficult to sequence and to clone in chicken, we supposed a bias could also exist for WGA. To check for correct amplification of microchromosomes, we designed markers from two scaffolds located on APL17, orthologous to GGA16 and containing the two major histocompatibility complex (MHC) gene clusters and the Nucleolar Organizing Region (NOR) rRNA genes, and from two scaffolds located on APL26, according to comparative genomic data given by the Narcisse database [31]. These 4 markers were added to our first set of 31 microsatellite markers we used primarily for selecting the 90 clones for the panel. Genotyping this set of markers on the WGA-DNA of the 90 hybrids demonstrated average retention frequencies of 23.8% for the whole genome, with 19.3% for the macrochromosomes and 29.9% for the microchromosomes. On average, retention frequencies are very close to those observed before the WGA (Figure 1A). However, retention frequency of S2870 located on APL17, S906 and S2549 located on APL26 were increased after WGA, especially for S2870. In contrast, a slight retention loss was found for S618, the other scaffold marker from APL17. Thus amplifying the panel by WGA and genotyping by the conventional PCR and agarose gel electrophoresis approach (WGA-PCR) appears to be a good option for mapping without having to perform large-scale culture of the hybrids, at least for macrochromosomes and medium-sized microchromosomes.

However, genotyping several thousand markers by individual PCR and gel electrophoresis would require a lot of time and effort and a higher throughput method would be more appropriate, if feasible. In addition to scaling down PCR volumes and reducing required DNA amounts, the Fluidigm BioMark™ has the added benefit of allowing rapid testing of 96 markers on 96 samples. To compare Fluidigm BioMark™ genotyping by qRT-

PCR, with (WGA-FLDMqPCR) or without (Pre-ampFLDMqPCR) WGA of the radiation hybrid DNA (see methods), with our more usual PCR and agarose (WGA-DNA) method, we used a set of 39 markers designed from scaffolds of the duck genome assembly. Results shown in Figure 1B suggest differences in retention frequencies between the three methods for the 39 markers tested, with lower values for the WGA-FLDMqPCR method. Differences in marker retention between the three methods was estimated by multiple t-tests (Table 1 and Additional file 3 Table S1), suggesting that there was no significant difference between the WGA-PCR and the Pre-ampFLDMqPCR methods, whereas the WGA-FLDMqPCR genotyping results were significantly different from the two others, with markedly lower retentions. These lower retentions values found with the WGA-FLDMqPCR condition are probably due to a lower sensitivity of the method, when compared to the Pre-ampFLDMqPCR condition (Figure 2). Taken together, our results suggest that our genotyping method by qPCR using the Fluidigm BioMark™ IFC Dynamic Array™ should not be performed using WGA DNA and also that if WGA DNA can be used for genotyping macrochromosome markers by the conventional agarose technique, it may cause problems for the smallest microchromosomes, as suggested by the results from the two markers on APL17.

To investigate further the possibility of using the Pre-ampFLDMqPCR method for genotyping, we constructed a map for a medium-size microchromosome (APL12), in addition to a first map of APL22 constructed by conventional PCR and agarose genotyping.

RH mapping of APL22 by WGA-PCR

Twenty-four scaffold markers derived from 15 duck scaffolds aligned to GGA21 in the Narcisse database (Figure 3) and designed as described in the Material and Methods section, were genotyped on WGA DNA by conventional PCR and gel electrophoresis. To build RH map of microchromosome APL22, two methods were used: one using the Carthagene software with the usual method [41] and a second using a comparative approach

based on the chicken genome, and the construction of robust map (see Methods). By the classical Carthagene approach, 24 markers were included in a single linkage group with a LOD score threshold of 11, and a framework map containing 12 markers and spanning 170 cR was obtained. However, five of the comprehensive map markers might extend the current map length by 53 cR and the most likely position for all framework and non-framework markers given by Carthagene software [41] are indicated in italics on the APL22 RH map (Figure 4). The comparative mapping approach and the associated robust map construction produced a map 283 cR long, containing 12 robust markers (Figure 4). The average retention frequency for the markers is 30.4%, in accordance with microchromosome retention frequency of the panel. A maximum marker retention around marker sca246B, suggests the centromere could be in that region (data not shown), which would be compatible with an acrocentric microchromosome. Comparative mapping with chicken chromosome GGA21 suggests several intrachromosomal rearrangements within this microchromosome.

RH mapping of APL12 by Pre-ampFLDMqPCR

Genotyping data for ten APL12 microsatellites and thirty-one markers designed from 18 scaffolds aligned to GGA11 were successfully obtained using the Pre-ampFLDMqPCR method and used to generate a RH map by the classical approach with the Carthagene software [41] and by the comparative mapping approach. After two-point analysis at a LOD threshold of 6, three linkage groups were defined among which the largest one contained 38 markers. The order of the 38 markers from the largest linkage group was determined by multipoint analysis with Carthagene and a framework map of APL12 bearing eighteen markers was obtained. The framework map is composed of 18 markers, covers 408 cR_{6000} and twenty additional markers on the comprehensive map extend the current map length by 34 cR (Figure 5). The map obtained by the comparative mapping approach is 728 cR long. The average retention for the markers on APL12 is 46%, significantly higher than the average

microchromosome retention. As a result, the whole chromosome has a relatively high retention rate and no position for the centromeric region can be suggested from the RH map.

Only one major intrachromosomal rearrangement is observed when comparing APL12 to chicken chromosome GGA11. One additional minor inversion is observed only when comparing GGA11 with the map of APL12 built with the classical Carthagene approach. The major inversion was tested and confirmed by FISH mapping using BAC clones located at both ends of the inverted fragment and corresponding to the regions of scaffold2558 and scaffold1176 (Figure 5). FISH results confirm the inversion (Figure 6).

RH Mapping of *no hit* EST markers from the smallest microchromosomes

Next, we wanted to test the three genotyping methods for mapping the smallest microchromosomes, orthologous to those absent from the current chicken sequence assembly and maps. We previously reported a strategy for mapping genes on the smallest microchromosomes absent from the chicken genome assembly [42, 43]. Chicken EST contigs with sequence similarity to the human genome and showing no BLAST hit in the chicken genome were selected to develop PCR markers. Most of these markers, which we named the *no hit* markers (see materials and methods), were found to cluster in specific regions of the human genome, likely corresponding to conserved syntenies missing in chicken and corresponding to the missing microchromosomes [42].

To increase chances of our markers showing linkage in duck, we focused marker development on duck EST contig sequence having sequence similarity to HSA19, in a region that was already shown to have synteny conservation with some of the smallest chicken chromosomes and being absent in the chicken genome assembly [43]. Due to the limited amount of duck EST data available, we were able to design only eight such markers derived

from duck EST contig data showing no significant BLAST hits with the current chicken assembly (chicken *no hit* markers) but with sequence similarity to human chromosome HSA19 (Figure 7). These were genotyped by all three techniques. Genotyping results for these eight *no hit* to chicken markers are showed Table 2, suggesting WGA has a much lower efficiency for the smallest microchromosomes, especially when used in combination with the FLDM method, leading to the underrepresentation (much lower retentions) or even to the total loss of the corresponding mapping data. For instance, the genotyping results of 8 *no hit* markers showed that some regions like the fragment spanned by marker *EstCtg23833* is not amplified by WGA because no positive signal was observed both in amplified hybrid DNA or in amplified duck genome DNA, whereas the remaining seven markers have a very low average retention: 5% in WGA-FLDMqPCR and 12% in WGA-PCR, compared to 34% for Pre-ampFLDMqPCR. The latter method seems thus the only one suited for mapping the smallest microchromosomes.

Analysis of the results with Carthagene showed that 4 out of the 8 markers: *EstCtg11412, EstCtg23833, YO3G5XE5* and *EstCtg293* are linked together and define a region of conserved synteny with HSA19 and GGA30 (Figure 7) and corresponds thus to APL31. The duck marker *EstCtg727* labels the gene *CKM* which is located very close to human genes *BCKDHA, SNRPA, MRPS12* and *PSMD8* which were shown to be on GGA32. This suggests that *EstCtg727* could be located on duck chromosome APL33.

Testing scaffold assembly

To test the quality of scaffold assembly, we selected the 13 largest duck scaffolds whose length ranged from 4.0 to 5.9 Mb and designed 70 markers with a density of one marker every megabase. These 70 markers were genotyped by Pre-ampFLDMqPCR and the results allowed the detection of one possible misassembly in scaffold504, for which a marker

located at one end showed no linkage with the others. Results for all the scaffolds are showed in Additional File 1 Figure S1.

To further test scaffolds from the duck genome assembly, we screened for potential chimeras by comparative mapping and detected using Narcisse [32], 41 duck scaffolds each of which mapping to two chicken chromosomes (Figure 8). As no inter-chromosomal rearrangements have been described to date between duck and chicken, we suspected assembly errors could have occurred and therefore tested 19 of the breakpoints by RH mapping with 45 markers. Results showed that all scaffolds, with the notable exception of sca649 could be misassembled (Figure 8 and Additional File 2 Figure S2).

DISCUSSION

Overall, the pattern of retention for the broken duck chromosome fragments in the hamster cells obtained here is very similar to that observed for the chicken radiation hybrid panel, with higher retentions for microchromosomes than for macrochromosomes. However, whereas only 23 % of the chicken-hamster hybrids produced had sufficient retention frequency values to be retained in the final panel, 50 % of the duck-hamster clones did. Indeed, although the fusion efficiency for chicken-hamster hybrids was reported to be as high as 2-9 x 10^{-6} by Kwok *et al* [44], it was only approximately 1.4 x 10^{-6} in our hands when we produced the 452 clones for the chicken whole genome RH panel. Here, the fusion efficiency is close to 3.5 x 10^{-6} which is three times higher. Such differences could be due to variations in chromosome structure and/or content between the two bird species or to differences in culture conditions. For instance, the *HPRT* gene used as a selection marker for the clones is on the short arm of macrochromosome GGA4 in chicken [45] and thus very likely to be on microchromosome APL10 in duck. Microchromosomes being better retained than the macrochromosomes, having the selection gene on one of them could help recovering a higher number of clones in each fusion experiment. It is also very likely that these results are due to

our change in culture conditions after the cell fusions: the chicken-hamster hybrids were cultivated at 40°C, the usual temperature for avian cells, whereas the duck-hamster ones at 37°C. Similarly, Ekker *et al* [46] succeeded in producing zebrafish somatic hybrids at 37°C but not at 28°C, which is the normal temperature for the culture of zebrafish cells. More generally, the difference in optimal temperatures for the growth of donor and recipient cells may be one of the possible causes for the lower retention frequencies usually observed for somatic and radiation hybrids in non mammalian species.

To obtain the DNA quantity required for building genome-wide maps, large-scale culture of the hybrid clones is necessary. However, in this process, donor DNA is lost. For instance, Karere *et al* [47] reported a genome wide half-life of the donor DNA of 8.7 passages and when preparing the whole genome RH (WGRH) panel in chicken, we observed the loss of 10% of the chicken genome after large cell culture of the hybrids [38]. This problem, in addition to the fact that large-scale culture of a RH panel requires lots of labor, encouraged us to find an alternative, such as WGA or scaling down the reaction volumes. Since the 1990s three major whole genome amplification techniques including primer extension pre-amplification (PEP) [48], degenerate oligonucleotide primed (DOP) PCR [49] and multiple displacement amplification (MDA) have been developed to address the problem of limiting amounts of DNA samples. PEP and DOP are both PCR-based methods and are limited by features of the Taq polymerase: typical amplification fragment length of < 3 kb and an error rate of 3×10^{-5}. These methods also suffer from incomplete coverage and uneven amplification of the genomic loci of several orders of magnitude, with $10^{-2} \sim 10^{-4}$ and $10^{-3} \sim 10^{-6}$ fold amplification biases for PEP and DOP-PCR methods, respectively [50]. MDA is an isothermal amplification employing the high fidelity Phi29 phage DNA polymerase for DNA synthesis and strand displacement [51]. The genome coverage is much improved, with an estimation of only 2.2 % missing after WGA by the MDA method in mammals [52]. Karere *et*

383 *al* [47] confirmed that MDA was suitable for RH mapping and reported a high concordance
384 rate of 97.6% with data from genomic DNA. However, even if this is true for mammals, it
385 might not be the case of microchromosomes in an avian genome.

386 When comparing retention frequencies before and after WGA in the 90 hybrids, with
387 the 35 markers used for clone selection, only slight variations of retention, either gains or
388 losses, were usually observed. However three markers, *CAUD064, S618* and *CAUD022,* show
389 an important loss of retention frequency after WGA while two others, *CAUD013* and *S2870,*
390 show a high increase, suggesting potential coverage problems by the WGA, either by lack of
391 coverage (losses) or by the over-representation of a region (gains). Moreover, genotyping of
392 eight *no hit* EST markers on WGA DNA, either using conventional PCR and Agarose or
393 FLDMqPCR, demonstrated a very low retention which is not in accordance with the retention
394 levels usually observed for microchromosomes. Therefore, we suggest that the genomic
395 features in the smallest microchromosomes causing coverage problems in whole genome
396 sequencing projects may also interfere with the efficiency of WGA. As we have already
397 shown, RH mapping can allow building maps for non-sequenced chromosomes [42, 43], it is
398 important that we produce genotyping results for them.

399 In this context, the Fluidigm BioMark™ IFC Dynamic Array™ genotyping method
400 can be an alternative to WGA, as only minute amounts of DNA (as little as 70 pg) are
401 required. High throughput gene expression analysis by real time PCR in a microfluidic
402 dynamic array was first introduced by Spurgeon *et al* [40], and has since been successfully
403 applied to copy number variation studies [53] and quantitative miRNA expression analysis
404 [54]. In our case, by performing qPCR with the Fluidigm BioMark™ IFC Dynamic Array™
405 genotyping, the additional benefit is high throughput, as the identification of bands on gel
406 electrophoresis is replaced by monitoring the PCR with Ct (Cycle threshold) and end point
407 Tm (melting temperature) values, allowing the distinction between specific and non-specific

amplification profiles. The Tm value is mainly influenced by base composition of amplicons, making it a specifically interesting parameter to follow when using markers defined from coding regions, which are more prone to cross-amplifying the hamster DNA.

We tested the Fluidigm genotyping method on WGA DNA and on standard DNA, with a pre-amplification step using a mix of all primers of the 96 markers analyzed together in a run [55]. In the WGA-FLDMqPCR runs, Ct values for the duck positive control was high with an average of 22 cycles (data not shown), as opposed to an average of 12 cycles, which is in the recommended scale, for the Pre-ampFLDMqPCR runs (Figure 3). These high Ct values suggested the quantity of DNA template was too low [55]. For a variety of reasons, WGA coupled with either FLDMqPCR or conventional PCR and agarose electrophoresis was unsuitable for genotyping on the smallest microchromosomes. Therefore, although the combination of WGA and FLDMqPCR would have allowed us to use less RH DNA, we decided the best genotyping method was to use standard DNA by FLDMqPCR genotyping, with a pre-amplification step performed using a mix of all primers for a set of 96 markers.

The drawbacks of genotyping by Fluidigm BioMark™ IFC Dynamic Array™ come from the fact that all 96 markers are genotyped with the same condition and therefore special care must be taken in the marker design. As a consequence, approximately 10 % of the markers were discarded during the final analysis due to poor quality data.

Apart from improving the genome assembly by assigning and ordering scaffolds to chromosomes, the duck RH panel can be used to test the scaffold assemblies. We tested this by genotyping markers at Mb density on the 13 scaffolds larger than 4 Mb, spanning altogether 60 Mb and thus accounting for 5.5% of the current duck genome assembly. A total of 70 markers were genotyped, only one marker (sca504F) on the end of *sca504* was not linked with other markers derived from the same scaffold (Additional File 1 Figure S1), suggesting an overall good quality of the final genome. To test further our capacity for

433 detecting potentially misassembled scaffolds, we took advantage of previously published data
434 indicating that on the whole, avian chromosomes are known to be well conserved throughout
435 evolution and more specifically, that no inter-chromosomal rearrangements, apart from the
436 well documented case of GGA4 = APL4 + APL10, have been discovered between chicken
437 and duck by current comparative cytogenetic approaches [17, 23, 28]. The 41 scaffolds
438 (including sca504) we detected as potential chimeras by comparative mapping had poor pair-
439 end sequence support (BGI, personal communication), suggesting most of them could indeed
440 be misassembled (Additional file 4 Table S2). We tested nineteen of them by genotyping
441 markers flanking the potential breakpoints (Additional File 2, Figure S2). As a result, all but
442 one scaffold (*sca649*) could be misassembled, and *sca649* possibly suggesting the first
443 detection of a small inter-chromosomal rearrangement between the duck and chicken
444 genomes, or perhaps more likely a segmental duplication in duck or in the last common
445 ancestor of the two species. This would need further confirmation by FISH mapping with
446 chicken BAC clones. It can be noted that the pair-end sequence support for this scaffold was
447 high, showing an agreement between sequencing and RH mapping data. When disagreements
448 between assembly and our RH data are detected in large scaffolds, they tend to happen
449 towards the end. To achieve better assembly accuracy, higher sequencing depth or more
450 efforts on developing sequencing libraries with longer inserts are needed.

451 Concerning the smallest duck microchromosomes, paralogous to those absent from the
452 chicken assembly, we suspect similar problems will arise: lack of sequence information,
453 difficulties in cloning, in genetic mapping, etc. RH mapping has proved useful for getting a
454 grip on these regions and one striking example is the case of some regions of HSA19, to
455 which no corresponding chicken genome data could be assigned by sequence similarity and to
456 which many chicken *no hit* EST showed significant sequence similarity. RH mapping with
457 these markers allowed building maps for GGA30 and GGA32 [43]. By developing markers

458 targeted to this region, a small linkage group composed of 4 *no hit* markers (absent in the
459 chicken genome assembly) orthologous to HSA19 was obtained. When aligned to HSA19, we
460 found they spanned a 5Mb region on HSA19p. Due to the lack of BAC clones for FISH or
461 other supplementary information, we cannot identify the duck chromosome, but according to
462 known data on synteny conservation between chicken and duck, we suggest that this small
463 linkage group should be assigned to APL31. Of the 8 *no hit* to chicken markers we studied
464 three have hits with small to medium-size scaffolds (between 23 and 96 kb) of the duck
465 assembly, suggesting that more sequence from the smallest microchromosomes could be
466 obtained in NGS (Table 2). Chicken/duck comparative mapping of GGA21/APL22 and
467 GGA11/APL12 microchromosomes demonstrate several intrachromosomal rearrangements,
468 the first described for microchromosomes in this pair of species. The maps obtained using the
469 usual Carthagene mapping approach or the comparative approach are very similar, apart for a
470 few markers, especially non-framework / non-robust ones, for which lower reliability in map
471 position can be due to the limits of the possible resolution of the mapping or to genotyping
472 errors. As the comparative approach starts with an ordering of markers corresponding to
473 chicken, it is interesting to note that the major duck-chicken rearrangements found with the
474 Carthagene approach are confirmed. A second advantage of the comparative mapping
475 approach and the associated construction of robust maps is that the number of robust markers
476 obtained is usually higher than the number of framework markers in the classical approach.
477 The major inversion found between GGA11 and APL12 is confirmed by FISH mapping, but
478 also by sequence alignment of duck scaffolds on the chicken assembly. Indeed, scaffold736,
479 whose integrity is demonstrated by RH mapping, with markers *sca736A* and *sca736B*
480 positioned close to one another at 153 and 154 cR on the CarthageneRH map and 402 and 441
481 cR on the ComparativeRH-Robust map, is separated in two locations when aligned to the
482 chicken sequence (Figure 5). Likewise, although a more complex pattern of events accounts

for the differences between GGA21 and APL22, one of them is supported by scaffold246, whose integrity is demonstrated by RH mapping with three markers on the robust map, each of which are positioned in different regions when aligned to the chicken sequence. Another is supported by markers sca871_1 and sca871_2, which are co-localized on the RH map and are 1.4 Mb apart in chicken (Figures 3 and 4). When comparing the turkey and chicken genomes, Zhang *et al*. also confirmed evidence for 20-27 major rearrangements between the two species and found one inversion between GGA11 and MGA13. However, they did not observe any rearrangement between GGA21 and MGA23. The mean estimated divergence time between chicken and turkey is 47 million years and 81 between chicken and duck [56]. A higher number of rearrangements are thus expected between the two latter pair of species. Only one major interchromosomal difference -with APL4 and APL10 corresponding to GGA4q and GGA4p respectively [21]- and very few intrachromosomal rearrangements have been reported between the chicken and the duck karyotypes [18-23]. The rearrangements observed with our data between GGA21 and APL22 seem more complex for example, Sca246B, Sca246C and Sca246D are split by Sca1885 in both RH maps. Likewise, Sca367B and Sca367C are split by Sca3327 in both RH maps, whereas they are adjacent in the chicken sequence, and Sca148 markers are widely split in the ComparativeRH map, while adjacent in the chicken. Further investigations and more precise maps using different techniques such as FISH or BAC contig maps will be needed to confirm these rearrangements. The increased resolution obtained by RH mapping as compared to the FISH mapping performed to date show that intrachromosomal rearrangements might happen on a finer scale than shown until now. This means that although the simple ordering of the duck scaffolds along the chicken genome by sequence similarity helps for chromosome assignment, the duck sequence thus obtained will be wrong whenever large or small-scale rearrangements will have happened between the two species. The whole duck assembled sequence will have to be ordered using the whole genome

RH map which will be constructed in our laboratory, in conjunction with other mapping methods, such as genetic and/or BAC contig physical maps, the latter allowing finer mapping and orientation of small scaffolds.

CONCLUSION

The chicken WGRH panel has been used to construct chromosome RH maps and helped in the genome assembly or the mapping of some of the smallest microchromosomes [42, 43, 57]. Similarly, the duck WGRH panel presented here will also be a major contribution to duck genomics. RH mapping can be a complementary approach to NGS by allowing the assignment of scaffolds to duck chromosomes and furthermore, detailed RH maps will allow a precise estimation of the intrachromosomal rearrangements that have occurred between chicken and duck.

Using the chicken genome as model and in combination with survey sequencing, the construction of dense RH maps of a less studied bird such as duck can be made. By taking advantage of the conservation of syntenies, optimal orders can be proposed [58, 59], thus maximizing the information obtained as first proposed by Hitte *et al.* [60, 61]. Indeed in duck, a dense RH map combined with scaffold sequencing and comparison to the chicken sequence, should lead to an improved genome assembly.

MATERIAL AND METHODS

Generation of radiation hybrids

The method was adapted from Morisson *et al* 2002 [38]. Normal diploid fibroblasts were obtained from 12-day-old Peking duck embryos from a highly inbred duck line. For each embryo, primary cells were obtained after trypsinization of the embryo tissues and the

rest of the tissues were stored for DNA extraction. Duplex PCR was performed to test the sex of embryos according to Batellier *et al* [62]. Fibroblasts from only one female embryo were propagated in complete DMEM medium (DMEM Glutamax (Gibco Co.) supplemented with 10% foetal calf serum (Gibco Co.), 1% penicillin and streptomycin) at 40°C with 5% CO_2 and used as donor cells. The *HPRT* (Hypoxanthine guanine phosphoribosyl transferase) -deficient hamster cell line Wg3hCl_2 [63] was used for recipient cells, which were cultured in complete RPMI medium (RPMI1640 (Sigma Chemical Co.) supplemented with 10% foetal calf serum (Gibco Co.), 1% penicillin and streptomycin) at 37°C with 5% CO_2. For each fusion experiment, 1.5×10^7 duck female fibroblasts were irradiated at 6000 rads by gamma rays from a Cesium-137 source and mixed to an equal number of Wg3hCl_2 hamster cells. The fusion partners were then pelleted and suspended in 1mL polyethyleneglycol (Roche Diagostics GmbH) and after 1min, 15mL DMEM medium without serum and antibiotics were gradually added from which 1 mL was taken to suspend in 10 mL complete RPMI medium and cultured at 37°C with 5% CO_2. Twenty-four hours after the fusion, HAT (hypoxanthine-aminopterin-thymidine) was added to the medium and four days later, the whole medium was changed to discard the non-fused cells. Eight to twelve days after the fusion, the first hybrid clones were observed. When fully grown after 7 to 20 days of culture, hybrids were picked and transferred to 25-cm^2 flasks. After confluence, the hybrid cells were subsequently transferred to two 75-cm^2 flasks. In order to limit the loss of duck fragment during the cell passages, hybrids were cultured for only one generation and harvested when fully grown. Ten million cells were kept for DNA extraction and the rest were cryoconserved.

Whole genome amplification

For each sample, 50ng starting RH DNA was amplified according to the GE Healthcare Illustra Genomiphi HY DNA amplification Kit protocol. To avoid representation bias, each hybrid was amplified in three replicates which were pooled to obtain the final

working panel DNA (WGA DNA). Duck genomic DNA, $Wg3hCl_2$ hamster DNA and H_2O were amplified in the same condition, as positive and negative controls. For genotyping experiments with the WGA DNA, the positive controls were duck genomic DNA and WGA duck genomic DNA, whereas when using standard DNA, both positive controls were duck genomic DNA.

In silico mapping of scaffolds to the chicken assembly, as a guide for choosing markers.

Seven thousand two hundred and five duck scaffolds larger than 1 kb were aligned to the current chicken assembly using Narcisse [31] and 1,787 were successfully positioned. All the data can be traced back at http://narcisse.toulouse.inra.fr/pre-narcisse/duck/cgi-bin/narcisse.cgi. According to Narcisse and existing comparative genomics data obtained by FISH [23, 28], approximate location of all the scaffold markers, especially chromosomal assignment could be inferred, but the real location still needed to be tested due to the possibilities of intrachromosomal rearrangement having occurred since chicken and duck divergence.

Thirty scaffolds were positioned on GGA11 and used for designing 31 potential APL12 markers, whereas 15 scaffolds were positioned on GGA21 from which 24 potential APL22 marker were derived.

Markers design

Altogether, 234 markers were used in our study and detailed information is given in Additional file 5 and Table S3. Twenty one microsatellite markers are from public databases (markers *APHXXX, CAUDXXX, AMUXXX and APTXXX*) and 10 *CAMXXX* markers were produced by our laboratory (Marie-Etancelin *et al.*, in preparation); 8 EST markers (*EstCtg11412, EstCtg23833, YO3G5XE5, EstCtg8099, Y04H5QR8, EstCtg2805***,** *EstCtg727*and *EstCtg293*) are from EST contig data (Pitel *et al*., Huang *et al*., in preparation);

the rest of the markers (ScaXXX or SXXX) were designed from the sequence of duck scaffolds from the genome assembly (Huang *et al*., in preparation) with the Primer3 software [64]. To avoid repetitive elements in the genome, the primers were checked by in-silico PCR [65] and the amplicon sequences aligned to the whole genome assembly by BLASTn.

Marker genotyping by conventional PCR and agarose gel electrophoresis (WGA-PCR.)

PCR reactions contained 25ng WGA DNA, 2mM $MgCl_2$, 0.5U Taq DNA polymerase (Promega Co.), 1X buffer (Promega), 200µM of each dNTP, 0.15µM of each forward primers, 0.2 µm of each reverse primers in a total volume of 15µL. PCR was performed on a GeneAmp PCR system 9700 thermocycler (Applied Biosystems): the first 5-min denaturation was followed by 48 cycles for microsatellite markers and 36 cycles for scaffold markers, each consisting of denaturation at 94 ^{o}C for 30s, annealing at specific temperature for 30s and elongation at 72 ^{o}C for 30s. PCR products were analyzed using a 2% agarose gel and were visualized by ethidium bromide staining. All the markers were genotyped in duplicate.

Marker genotyping by Fluidigm BioMark™ IFC Dynamic Array™ quantitative PCR on WGA DNA (WGA-FLDMqPCR)

WGA DNA (90 panel samples, positive control: standard Duck DNA and WGA Duck DNA, negative control: standard Hamster DNA and WGA Hamster DNA, blank control: WGA H_2O and H_2O) and an assay set containing 96 primer pairs in which concentration of each primer pair is 20 µM were loaded on a Fluidigm BioMark™ 96.96 Dynamic Array™ IFC. WGA DNA was quantified by Picogreen, the ideal concentration of the DNA was of 50 ng/µL. In fact, WGA DNA proved difficult to quantify by Picogreen, likely due to the complex branched structure of the amplification product obtained. Real time PCR was performed in the presence of EvaGreen™ DNA-binding dye, according to the manufacturer's protocol [40]. All the markers were genotyped in duplicate.

Marker genotyping by Fluidigm BioMark™ IFC Dynamic Array™ quantitative PCR on pre-amplified standard DNA (Pre-ampFLDMqPCR)

Standard DNA was quantified by Picogreen and diluted at a final concentration of 5 ng/µL. Primer pairs for the 96 markers included in one Fluidigm BioMark™ run were diluted at a final concentration of 20 µM and distributed in a 96-well microplate called a 20µM assay set. Then 8 µl of 0.1 M TE and 2 µl of each primer mix from the 20µM set were pooled in a 1 mL Eppendorf tube and vortexed thoroughly (96 Markers Primer mix). Pre-amplification was performed in 5 µL, containing 2.5 µl Pre-amplification Master mix (Applied Biosystems), 1.25 µl of 96 Markers Primer mix and 1.25 µl DNA at 5 ng/µL (90 panel samples, positive control: genomic Duck DNA, negative control: genomic Hamster DNA, blank control: H_2O). After denaturation for 10min at 95°C, a PCR was performed by 14 cycles of 15 s at 95 °C and 4min at 60°C, and a final elongation step at 20 °C for 10min. The pre-amplification products thus obtained were diluted 7 times before the Fluidigm BioMark™ run. The 96 diluted pre-amplified samples and 20 µM 96 primer pairs assay set was loaded on a Fluidigm BioMark™ 96.96 Dynamic Array™ IFC, using the same procedure as for the WGA-FLDMqPCR marker genotyping method. All the markers were genotyped in duplicate.

Interpretation of FLDM data

Data was analyzed using the Fluidigm Real-Time PCR Analysis software to obtain the Ct values (Cycle Threshold: number of cycles required for the fluorescent signal to cross a given threshold) and Tm values (DNA melting temperature which is influenced by the length and base composition of the DNA molecules amplified) (Figure 3). For the genotyping calling, the positive control (duck DNA) should not be too low or too high (Ct values between 10 and 16). A hybrid was called positive when the hybrid had a Ct value lower or equal to that of the negative control and a Tm value close to the positive control. A genotype was called "Unknown" when a hybrid had a high Ct value but the same Tm as the positive control or a

low Ct value but a Tm value was slightly different (± 1.5^{o}C) than the positive control. If no amplification of the positive control could be seen, the marker was discarded altogether.

Map construction

Two methods for map construction were used: (i) a classical approach by two point and multipoint mapping, followed by the determination of the minimal set of markers for a framework map and (ii) a comparative map approach with statistical measure of a set of maps. The classical RH map were constructed using the Carthagene software [41] in three steps: (1) linkage groups were defined by two point analysis using a LOD score threshold of 11 (for the RH map of APL22) or 6 (for the RH map of ALP12) (2) multipoint analyses were done to define a framework map for the larger linkage groups, using a LOD threshold of 3 for the framework maps (3) a comprehensive map was built by calculating the location of additional markers relative to the framework markers. The comparative map approach is described by Faraut *et al*., (2007) [58]. It uses the information of marker adjacencies in a related genome, to assist the mapping process when the experimental data is not conclusive, thus directly producing comparative maps minimizing the number of breakpoints. The comparative mapping is followed by a statistical confidence measure of a distribution of maps to evaluate map uncertainties and produce a robust map, such as described in Servin *et al*. (2010) [66]. Finally the map figures were created using MapChart [67].

Fluorescent in situ hybridisation (FISH).

Chicken BAC clones were chosen in the Wageningen BAC library [68] according to their known position, as estimated by BAC end sequence information, in regions paralogous to the breakpoint under study. WAG19G7 (accession number CZ566048) corresponds to the duck scaffold sca2558, while WAG13P2 (CZ561694) and WAG20C21 (CZ565661) correspond to sca1176. BAC clones were grown in LB medium with 12,5 µg/ml chloramphenicol. The DNA was extracted using the Qiagen plasmid midi kit.

FISH was carried out on metaphase spreads obtained from fibroblast cultures of 7-days old chicken and duck embryos, arrested with 0.05 µg/ml colcemid (Sigma) and fixed by standard procedures. The FISH protocol is derived from Yerle *et al*, 1992 [69]. Two-colour FISH was performed by labelling 100 ng for each BAC clones with alexa fluorochromes (ChromaTide® Alexa Fluor® 488-5-dUTP, Molecular probes; ChromaTide® Alexa Fluor® 568-5-dUTP, Molecular Probes) by random priming using the Bioprim Kit (Invitrogen). The probes were purified using spin column G50 Illustra (Amersham Biosciences). Probes were ethanol precipitated, resuspended in 50% formamide hybridization buffer (for FISH on chicken metaphases) or in 40% formamide hybridization buffer for heterologous FISH. Probes were hybridised to chicken metaphase slides for 17 hours at 37°C and to duck metaphases for 48H in the Hybridizer (Dako). Chromosomes were counterstained with DAPI in antifade solution (Vectashield with DAP, Vector). The hybridised metaphases were screened with a Zeiss fluorescence microscope and a minimum of twenty spreads was analysed for each experiment. Spot-bearing metaphases were captured and analysed with a cooled CCD camera using Cytovision software (Applied Imaging).

AUTHORS' CONTRIBUTIONS

MR, SB and MM produced the RH panel. MR made the whole genome amplification, designed some markers, made and analyzed the genotyping. KF and EL assisted with marker genotyping and genotyping analysis. YH and NL provided the duck sequencing and assembly data. AV and MR built the RH map. VF performed the FISH mapping. TF made the alignment of the duck scaffolds to the current chicken assembly. MR, AV and MM drafted the manuscript. TF, AV and MM conceived the study, participated in its coordination and finalized the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGEMENTS

This research was supported by a grant from the INRA Animal Genetics Division. MR was supported by the China scholarship council and the INRA Animal Genetics Division. The duck fibroblasts were irradiated at Plateforme Anexpo of the Toulouse Midi-Pyrenees Genopole (http://anexplo.genotoul.fr). Genotyping were performed at the Plateforme Génomique of the Toulouse Midi-Pyrenees Genopole (http://genomique.genotoul.fr/). We are very grateful to Chantal Delcros for technical assistance, to Dr Magali SanCristobal for providing advice for statistical tests and to Dr Jacqueline Gelfy (UMR INRA/ENVT Host - Pathogen Agent Interactions, Toulouse, France) for providing advice for duck fibroblast cultures. EST contig data produced by the INRA PYRESAVI project was kindly provided by SIGENAE.

COMPETING INTERESTS

The authors declare that they have no competing interests.

REFERENCES

1. Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, Li Y, Li S, Shan G, Kristiansen K *et al*: **De novo assembly of human genomes with massively parallel short read sequencing**. *Genome Res* 2010, **20**(2):265-272.
2. Li R, Fan W, Tian G, Zhu H, He L, Cai J, Huang Q, Cai Q, Li B, Bai Y *et al*: **The sequence and de novo assembly of the giant panda genome**. *Nature* 2010, **463**(7279):311-317.
3. Xia Q, Guo Y, Zhang Z, Li D, Xuan Z, Li Z, Dai F, Li Y, Cheng D, Li R *et al*: **Complete resequencing of 40 genomes reveals domestication events and genes in silkworm (Bombyx)**. *Science* 2009, **326**(5951):433-436.
4. Huang S, Li R, Zhang Z, Li L, Gu X, Fan W, Lucas WJ, Wang X, Xie B, Ni P *et al*: **The genome of the cucumber, Cucumis sativus L**. *Nat Genet* 2009, **41**(12):1275-1281.
5. Rubin CJ, Zody MC, Eriksson J, Meadows JR, Sherwood E, Webster MT, Jiang L, Ingman M, Sharpe T, Ka S *et al*: **Whole-genome resequencing reveals loci under selection during chicken domestication**. *Nature* 2010, **464**(7288):587-591.
6. Ye L, Hillier LW, Minx P, Thane N, Locke DP, Martin JC, Chen L, Mitreva M, Miller JR, Haub KV *et al*: **A vertebrate case study of the quality of assemblies derived from next-generation sequences**. *Genome Biol* 2011, **12**(3):R31.
7. Dalloul RA, Long JA, Zimin AV, Aslam L, Beal K, Blomberg Le A, Bouffard P, Burt DW, Crasta O, Crooijmans RP *et al*: **Multi-platform next-generation sequencing of the domestic turkey (Meleagris gallopavo): genome assembly and analysis**. *PLoS Biol* 2010, **8**(9).
8. Tuyen D: **The situation of duck production in Vietnam**. *Proceedings of the International Seminar on Improved Duck Production of Small-scale Farmers in the ASPAC region* 2007(Hanoi 2007):123-133.
9. Cheng Y: **Breeding and genetics of waterfowl**. *Worlds Poult Sci J* 2003(59):509-519.
10. Marie-Etancelin C. CH, Brun J.M., Marzul C., Mialon-Richard M.M., and Rouvier R.: **Genetics and selection of mule ducks in France: a review**. *Worlds Poult Sci J* 2008(64):187-207.
11. Sturm-Ramirez KM, Hulse-Post DJ, Govorkova EA, Humberd J, Seiler P, Puthavathana P, Buranathai C, Nguyen TD, Chaisingh A, Long HT *et al*: **Are ducks contributing to the endemicity of highly pathogenic H5N1 influenza virus in Asia?** *J Virol* 2005, **79**(17):11269-11279.
12. Olsen B, Munster VJ, Wallensten A, Waldenstrom J, Osterhaus AD, Fouchier RA: **Global patterns of influenza a virus in wild birds**. *Science* 2006, **312**(5772):384-388.
13. Gaidet N, Cattoli G, Hammoumi S, Newman SH, Hagemeijer W, Takekawa JY, Cappelle J, Dodman T, Joannis T, Gil P *et al*: **Evidence of infection by H5N2 highly pathogenic avian influenza viruses in healthy wild waterfowl**. *PLoS Pathog* 2008, **4**(8):e1000127.
14. van Asseldonk M.A.P.M. MMPM, Mourits M.C.M. and Huirne R.B.M.: **Economics of controlling avian influenza epidemics**. *In Avian Influenza Prevention and Control, Springer Ed Series: Wageningen UR Frontis Series* 2008, **8**:139-148.
15. Kim JK NN, Forrest HL, Webster RG: **Ducks: the "Trojan horses" of H5N1 influenza.** *Influenza Other Respi Viruses* 2009, **3**(4):**121-128**.
16. Burt DW: **Origin and evolution of avian microchromosomes**. *Cytogenet Genome Res* 2002, **96**(1-4):97-112.

17. Huang Y, Zhao Y, Haley CS, Hu S, Hao J, Wu C, Li N: **A genetic and cytogenetic map for the duck (Anas platyrhynchos)**. *Genetics* 2006, **173**(1):287-296.
18. Guttenbach M, Nanda I, Feichtinger W, Masabanda JS, Griffin DK, Schmid M: **Comparative chromosome painting of chicken autosomal paints 1-9 in nine different bird species**. *Cytogenet Genome Res* 2003, **103**(1-2):173-184.
19. Derjusheva S, Kurganova A, Habermann F, Gaginskaya E: **High chromosome conservation detected by comparative chromosome painting in chicken, pigeon and passerine birds**. *Chromosome Res* 2004, **12**(7):715-723.
20. Kayang BB, Fillon V, Inoue-Murayama M, Miwa M, Leroux S, Feve K, Monvoisin JL, Pitel F, Vignoles M, Mouilhayrat C *et al*: **Integrated maps in quail (Coturnix japonica) confirm the high degree of synteny conservation with chicken (Gallus gallus) despite 35 million years of divergence**. *BMC Genomics* 2006, **7**:101.
21. Griffin DK, Robertson LB, Tempest HG, Skinner BM: **The evolution of the avian genome as revealed by comparative molecular cytogenetics**. *Cytogenet Genome Res* 2007, **117**(1-4):64-77.
22. Griffin DK, Robertson LB, Tempest HG, Vignal A, Fillon V, Crooijmans RP, Groenen MA, Deryusheva S, Gaginskaya E, Carre W *et al*: **Whole genome comparative studies between chicken and turkey and their implications for avian genome evolution**. *BMC Genomics* 2008, **9**:168.
23. Skinner BM, Robertson LB, Tempest HG, Langley EJ, Ioannou D, Fowler KE, Crooijmans RP, Hall AD, Griffin DK, Volker M: **Comparative genomics in chicken and Pekin duck using FISH mapping and microarray analysis**. *BMC Genomics* 2009, **10**:357.
24. Warren WC, Clayton DF, Ellegren H, Arnold AP, Hillier LW, Kunstner A, Searle S, White S, Vilella AJ, Fairley S *et al*: **The genome of a songbird**. *Nature* 2010, **464**(7289):757-762.
25. Volker M, Backstrom N, Skinner BM, Langley EJ, Bunzey SK, Ellegren H, Griffin DK: **Copy number variation, chromosome rearrangement, and their association with recombination during avian evolution**. *Genome Res* 2010, **20**(4):503-511.
26. Zhang Y, Zhang X, O'Hare TH, Payne WS, Dong JJ, Scheuring CF, Zhang M, Huang JJ, Lee MK, Delany ME *et al*: **A comparative physical map reveals the pattern of chromosomal evolution between the turkey (Meleagris gallopavo) and chicken (Gallus gallus) genomes**. *BMC Genomics* 2011, **12**:447.
27. Hedges SB, Dudley J, Kumar S: **TimeTree: a public knowledge-base of divergence times among organisms**. *Bioinformatics* 2006, **22**(23):2971-2972.
28. Fillon V, Vignoles M, Crooijmans RP, Groenen MA, Zoorob R, Vignal A: **FISH mapping of 57 BAC clones reveals strong conservation of synteny between Galliformes and Anseriformes**. *Anim Genet* 2007, **38**(3):303-307.
29. Keightley PD, Eyre-Walker A: **Deleterious mutations and the evolution of sex**. *Science* 2000, **290**(5490):331-333.
30. van Tuinen M, Hedges SB: **Calibration of avian molecular clocks**. *Mol Biol Evol* 2001, **18**(2):206-213.
31. Courcelle E, Beausse Y, Letort S, Stahl O, Fremez R, Ngom-Bru C, Gouzy J, Faraut T: **Narcisse: a mirror view of conserved syntenies**. *Nucleic Acids Res* 2008, **36**(Database issue):D485-490.
32. **Narcisse Duck** [http://narcisse.toulouse.inra.fr/pre-narcisse/duck/cgi-bin/narcisse.cgi]
33. Kraus RH, Kerstens HH, Van Hooft P, Crooijmans RP, Van Der Poel JJ, Elmberg J, Vignal A, Huang Y, Li N, Prins HH *et al*: **Genome wide SNP discovery, analysis and evaluation in mallard (Anas platyrhynchos)**. *BMC Genomics* 2011, **12**:150.

34. Kraus RH, Kerstens HH, van Hooft P, Megens HJ, Elmberg J, Tsvey A, Sartakov D, Soloviev SA, Crooijmans RP, Groenen MA *et al*: **Widespread horizontal genomic exchange does not erode species barriers among sympatric ducks**. *BMC Evol Biol* 2012, **12**:45.
35. Yuan X, Zhang M, Ruan W, Song C, Ren L, Guo Y, Hu X, Li N: **Construction and characterization of a duck bacterial artificial chromosome library**. *Anim Genet* 2006, **37**(6):599-600.
36. Lawrence S, Morton NE, Cox DR: **Radiation hybrid mapping**. *Proc Natl Acad Sci U S A* 1991, **88**(17):7477-7480.
37. James MR, Richard CW, 3rd, Schott JJ, Yousry C, Clark K, Bell J, Terwilliger JD, Hazan J, Dubay C, Vignal A *et al*: **A radiation hybrid map of 506 STS markers spanning human chromosome 11**. *Nat Genet* 1994, **8**(1):70-76.
38. Morisson M, Lemiere A, Bosc S, Galan M, Plisson-Petit F, Pinton P, Delcros C, Feve K, Pitel F, Fillon V *et al*: **ChickRH6: a chicken whole-genome radiation hybrid panel**. *Genet Sel Evol* 2002, **34**(4):521-533.
39. Barrett JH: **Genetic mapping based on radiation hybrid data**. *Genomics* 1992, **13**(1):95-103.
40. Spurgeon SL, Jones RC, Ramakrishnan R: **High throughput gene expression measurement with real time PCR in a microfluidic dynamic array**. *PLoS One* 2008, **3**(2):e1662.
41. de Givry S, Bouchez M, Chabrier P, Milan D, Schiex T: **CARHTA GENE: multipopulation integrated genetic and radiation hybrid mapping**. *Bioinformatics* 2005, **21**(8):1703-1704.
42. Douaud M, Feve K, Gerus M, Fillon V, Bardes S, Gourichon D, Dawson DA, Hanotte O, Burke T, Vignoles F *et al*: **Addition of the microchromosome GGA25 to the chicken genome sequence assembly through radiation hybrid and genetic mapping**. *BMC Genomics* 2008, **9**:129.
43. Morisson M, Denis M, Milan D, Klopp C, Leroux S, Bardes S, Pitel F, Vignoles F, Gerus M, Fillon V *et al*: **The chicken RH map: current state of progress and microchromosome mapping**. *Cytogenet Genome Res* 2007, **117**(1-4):14-21.
44. Kwok C, Korn RM, Davis ME, Burt DW, Critcher R, McCarthy L, Paw BH, Zon LI, Goodfellow PN, Schmitt K: **Characterization of whole genome radiation hybrid mapping resources for non-mammalian vertebrates**. *Nucleic Acids Res* 1998, **26**(15):3562-3566.
45. Fukagawa T, Hayward N, Yang J, Azzalin C, Griffin D, Stewart AF, Brown W: **The chicken HPRT gene: a counter selectable marker for the DT40 cell line**. *Nucleic Acids Res* 1999, **27**(9):1966-1969.
46. Ekker M, Speevak MD, Martin CC, Joly L, Giroux G, Chevrette M: **Stable transfer of zebrafish chromosome segments into mouse cells**. *Genomics* 1996, **33**(1):57-64.
47. Karere GM, Lyons LA, Froenicke L: **Enhancing radiation hybrid mapping through whole genome amplification**. *Hereditas* 2010, **147**(2):103-112.
48. Zhang L, Cui X, Schmitt K, Hubert R, Navidi W, Arnheim N: **Whole genome amplification from a single cell: implications for genetic analysis**. *Proc Natl Acad Sci U S A* 1992, **89**(13):5847-5851.
49. Telenius H, Carter NP, Bebb CE, Nordenskjold M, Ponder BA, Tunnacliffe A: **Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer**. *Genomics* 1992, **13**(3):718-725.
50. Silander K, Saarela J: **Whole genome amplification with Phi29 DNA polymerase to enable genetic or genomic analysis of samples of low DNA yield**. *Methods Mol Biol* 2008, **439**:1-18.

51. Dean FB, Hosono S, Fang L, Wu X, Faruqi AF, Bray-Ward P, Sun Z, Zong Q, Du Y, Du J *et al*: **Comprehensive human genome amplification using multiple displacement amplification**. *Proc Natl Acad Sci U S A* 2002, **99**(8):5261-5266.
52. Lasken RS: **Genomic DNA amplification by the multiple displacement amplification (MDA) method**. *Biochem Soc Trans* 2009, **37**(Pt 2):450-453.
53. Qin J, Jones RC, Ramakrishnan R: **Studying copy number variations using a nanofluidic platform**. *Nucleic Acids Res* 2008, **36**(18):e116.
54. Jang JS, Simon VA, Feddersen RM, Rakhshan F, Schultz DA, Zschunke MA, Lingle WL, Kolbert CP, Jen J: **Quantitative miRNA expression analysis using fluidigm microfluidics dynamic arrays**. *BMC Genomics* 2011, **12**:144.
55. Mengual L, Burset M, Marin-Aguilera M, Ribal MJ, Alcaraz A: **Multiplex preamplification of specific cDNA targets prior to gene expression analysis by TaqMan Arrays**. *BMC Res Notes* 2008, **1**:21.
56. **Time Tree** [http://www.timetree.org/index.php]
57. Solinhac R, Leroux S, Galkina S, Chazara O, Feve K, Vignoles F, Morisson M, Derjusheva S, Bed'hom B, Vignal A *et al*: **Integrative mapping analysis of chicken microchromosome 16 organization**. *BMC Genomics* 2011, **11**:616.
58. Faraut T, de Givry S, Chabrier P, Derrien T, Galibert F, Hitte C, Schiex T: **A comparative genome approach to marker ordering**. *Bioinformatics* 2007, **23**(2):e50-56.
59. Faraut T, de Givry S, Hitte C, Lahbib-Mansais Y, Morisson M, Milan D, Schiex T, Servin B, Vignal A, Galibert F *et al*: **Contribution of radiation hybrids to genome mapping in domestic animals**. *Cytogenet Genome Res* 2009, **126**(1-2):21-33.
60. Hitte C, Madeoy J, Kirkness EF, Priat C, Lorentzen TD, Senger F, Thomas D, Derrien T, Ramirez C, Scott C *et al*: **Facilitating genome navigation: survey sequencing and dense radiation-hybrid gene mapping**. *Nat Rev Genet* 2005, **6**(8):643-648.
61. Hitte C, Kirkness EF, Ostrander EA, Galibert F: **Survey sequencing and radiation hybrid mapping to construct comparative maps**. *Methods Mol Biol* 2008, **422**:65-77.
62. Batellier F, Marchal F, Scheller MF, Gautron J, Sellier N, Taouis M, Monbrun C, Vignal A, Brillard JP: **Sex ratios in mule duck embryos at various stages of incubation**. *Theriogenology* 2004, **61**(2-3):573-580.
63. Echard G. GJ, Gillois M.: **Localisation des gène MPI , PKM2,NP sur le chromosome3 de porc (Sus scrofa L.) et analyse cytogénétique d'une lignée de hamster chinois issuede la DON (Wg3h)**. *Genet Sel Evol* 1984, **16**:261-270.
64. **Primer3** [http://frodo.wi.mit.edu/]
65. **UCSC In-Silico PCR** [http://genome.csdb.cn/cgi-bin/hgPcr]
66. Servin B, de Givry S, Faraut T: **Statistical confidence measures for genome maps: application to the validation of genome assemblies**. *Bioinformatics* 2010, **26**(24):3035-3042.
67. Voorrips RE: **MapChart: software for the graphical presentation of linkage maps and QTLs**. *J Hered* 2002, **93**(1):77-78.
68. Crooijmans RP, Vrebalov J, Dijkhof RJ, van der Poel JJ, Groenen MA: **Two-dimensional screening of the Wageningen chicken BAC library**. *Mamm Genome* 2000, **11**(5):360-363.
69. Yerle M, Galman O, Lahbib-Mansais Y, Gellin J: **Localization of the pig luteinizing hormone/choriogonadotropin receptor gene (LHCGR) by radioactive and nonradioactive in situ hybridization**. *Cytogenetics and Cell Genetics* 1992, **59**:48-51.

FIGURES and TABLES

Figure 1: Estimations of duck genome retention in the RH clones. A: retention frequencies of thirty-one microsatellite markers and four scaffold markers before (white) and after (grey) whole genome amplification. The test was done on the 90 selected hybrids by conventional Agarose genotyping. The expected chromosome locations of the markers (given in brackets) are derived from the chicken/duck comparative FISH mapping and a duck genetic map (Marie-Etancelin *et al.*, in prep) for the microsatellite markers and according to comparative genomic data given by the Narcisse software [32] for the scaffold markers.

B: Retention frequencies of thirty-nine scaffolds markers obtained using three different genotyping strategies. The thirty-nine scaffold markers were genotyped using either (i) the amplified panel with conventional agarose genotyping (blue: WGA-PCR), (ii) the non amplified panel and genotyping with the Fluidigm BioMark gene expression dynamic array (green: Pre-ampFLDMqPCR) or (iii) the amplified panel and genotyping with the Fluidigm BioMark™ IFC Dynamic Array™ genotyping by quantitative PCR without any pre-amplification step (purple: WGA-FLDMqPCR). The markers are distributed along the X axis from the lowest to the highest retention frequencies obtained by the first method (the amplified panel with conventional agarose genotyping WGA-PCR in blue).

Figure 2: Genotyping by Fluidigm BioMarkTM IFC Dynamic ArrayTM quantitative PCR. (A) WGA-FLDMqPCR: WGA-amplified DNA and qPCR. Left: double-strand DNA (dsDNA) accumulation curve as a function of the number of cycles. Right: melting curve of the final product. Green: positive control (duck DNA). Red: a hybrid which was positive (containing duck DNA corresponding to the marker tested). Blue: a negative hybrid. Yellow: negative control (hamster DNA). (B) Pre-ampFLDMqPCR: non-amplified DNA, a pre-amplification step with a mix of the 96 primer pairs for the 96 markers tested in the Fluidigm BioMarkTM assay and qPCR. The same markers and controls are used as in (A). The

sensitivity is higher in (B), with a lower number of cycles necessary for detection of duck DNA. The negative control and the hybrid not containing duck DNA amplify at a much higher number of cycles and the non-specific products amplified can easily be distinguished by their different melting temperature values (right). In both experiments, no amplification was obtained from water (data not shown)**.**

Figure 3: Developing markers using comparative mapping data. Screenshot of GGA21 from the Narcisse database (http://narcisse.toulouse.inra.fr/pre-narcisse/duck/cgi-bin/narcisse.cgi). Right: GGA21, with gene names. Left: white cylinders represent duck scaffolds or portions of duck scaffolds aligned to the chicken genome. Grey and green arrows represent portions of conserved synteny between the chicken chromosome and the duck scaffolds and their orientation. Left: names of the markers developed for RH mapping. For large scaffolds, such as sca148, one marker every 500 kb was developed to ensure RH linkage by optimizing inter-marker distances. Red: scaffold246 and green: scaffold871. These two scaffolds each seem to be split in chicken into three and two different regions respectively. At least one marker per region was developed, so as to check duck scaffold integrity.

Figure 4: **Comparative mapping between chicken chromosome 21 (GGA21) sequence map and duck chromosome 22 (APL22) radiation hybrid maps.** Left and right: position of duck scaffold markers on the chicken genome. Middle left: RH map built with the Carthagene software. Middle right: RH map built with the comparative approach, followed by statistical confidence measures for genome maps. Framework markers for the CarthageneRH map and robust markers for the ComparativeRH map are in red.

Figure 5: **Comparative mapping between chicken chromosome 11 (GGA11) sequence map and duck chromosome 12 (APL12) radiation hybrid maps.** Left and right: position of duck scaffold markers on the chicken genome. Middle left: RH map built with the Carthagene

947 software. Middle right: RH map built with the comparative approach, followed by statistical
948 confidence measures for genome maps. Framework markers for the CarthageneRH map and
949 robust markers for the ComparativeRH map are in red or in blue (inversion). Two markers
950 boxed in red correspond to the chicken BAC clones used for FISH mapping.
951 **Figure 6: Confirmation of the inversion on APL12 by FISH.** Chromosomes are stained by
952 DAPI**.** Centromere positions (cen) are indicated by arrows. Left: BAC clone WAG19G7,
953 corresponding to duck scaffold sca2558 is located in the centromeric region of GGA11 (top),
954 whereas it is clearly located in the middle of the q arm of APL12 (bottom), suggesting the
955 occurrence of an intrachromosomal rearrangement. Right: BAC clone WAG19G7 (red)
956 corresponds to scaffold2558, whereas WAG20C21 and WAG13P2 (green) to scaffold1176. In
957 chicken WAG19G7 (scaffold2558) is located in the centromeric region of GGA11 and
958 WAG20C21 (scaffold1176) is in the middle of the q arm (top), whereas in duck, WAG19G7
959 (scaffold2558) is located in the middle of the q arm and WAG13P2 (scaffold1176) is at the
960 end (bottom). This suggests the occurrence of an inversion between the two species. The
961 black bands in the middle of APL12 near BAC clone 19G7, might be an artifact resulting
962 from over-denaturation or to the DAPI staining.

963

964 **Figure 7: Chicken and duck microchromosome linkage groups based on 'no hit' EST**
965 **mapping.**
966 -*Left* : the chicken linkage groups are from Morisson *et al.*, 2007 [43]. Markers were
967 developed from chicken EST contigs absent from the chicken assembly (*no hit* markers),
968 presenting sequence similarity to HSA19. Markers in blue, purple or green are '*no hit*' EST;
969 genetic markers are in red and framework markers are underlined. Markers in black got
970 subsequently included in the linkage groups.

-*Middle:* position on HSA19 of chicken EST markers (blue, purple or green) and duck EST markers (brown). For each marker, the name of the gene is added. The duck EST markers are shown on both sides of the map to allow visualization of all possible pair wise map comparisons.

-*Right:* a duck RH linkage group corresponding to one part of chicken microchromosome GGA30. They both bear the genes AKAP8 (GCT1867 in chicken and EstCtg293 in duck) and KEAP1 (GCT1859 in chicken and Y03G5XE5, EstCtg23833 in duck). Markers were developed from duck EST contigs, presenting sequence similarity to HSA19 and for which no sequence similarity could be found on the chicken genome.

Figure 8: Testing duck scaffolds aligning to two chicken chromosomes. Based on previous observations, duck scaffolds aligning to two chicken chromosomes were suspected to be misassembled and one example is shown here. A: sca3008, boxed in red, aligns to GGA5 and GGA7, according to the Narcisse database. B: Markers sca3008A (green) and sca3008B (purple), very close to one another on sca3008, but spanning the putative breakpoint, were genotyped on the RH panel, but failed to show linkage, indicating that the scaffold is indeed misassembled. Results for other scaffolds are shown in Additional File 2 Figure S2.

	WGA-PCR	WGA-FLDMqPCR	Pre-ampFLDMqPCR
WGA-PCR	**26.1**	15.9	24.8
WGA-FLDMqPCR	7.4e-08	**16.2**	15.8
Pre-ampFLDMqPCR	0.7	2.1e-10	**28.1**

Table 1: Comparison of marker retention with the three genotyping techniques. Diagonal (in bold): mean number of positive hybrids in the panel (90 hybrids; 39 markers tested). Above the diagonal: mean number of positive hybrids in common between two conditions. Below the diagonal: P-values adjusted by Bonferroni correction for the differences in marker retention between two techniques.

	WGA-PCR			WGA-FLDMqPCR			Pre-ampFLDMqPCR		BLAST to Duck Assembly	
	WGA Duck[1]	Duck[2]	No. pos[3]	WGA Duck[1]	Duck[2]	No. pos[3]	Duck[2]	No. pos[3]	BLAST hit (scaffold name)	Scaffold length (bp)
EstCtg11412	+	+	11	+	+	3	+	25	sca4924	26 914
EstCtg23833	-	+	0	-	+	0	+	25	C19155564	548
EstCtg2805	+	+	18	-	+	3	+	24	sca12946	245
EstCtg293	+	+	24	+	+	16	+	30	sca271	23 394
EstCtg727	+	+	14	-	+	1	+	44	nohit	NA
EstCtg8099	-	+	1	-	+	2	+	29	C18154597	159
Y03G5XE5	+	+	7	-	+	0	+	25	nohit	NA
04H5QRB	+	+	13	+	+	11	+	43	sca1017	95 902
Nb. Controls[4] or Mean pos[5]	6/8	8/8	11	3/8	8/8	4.5	16/16	30.6	NA	NA
Mean retentions (%)	NA	NA	12	NA	NA	5	NA	34	NA	NA

Table 2: **Genotyping 8 *no hit* markers using three different genotyping strategies.** The 8 *no hit* markers were genotyped using either of three methods (see Material and Methods): (i) WGA-PCR: the WGA-amplified panel and conventional agarose genotyping; (ii) WGA-FLDMqPCR: the WGA-amplified panel and genotyping with the Fluidigm BioMark gene expression dynamic array, without the pre-amplification with a mix

of all primer pairs or (iii) Pre-ampFLDMqPCR: the non-amplified panel and genotyping with the Fluidigm BioMark™ IFC Dynamic Array™ with a pre-amplification step using a mix of all primer pairs. [1]WGA Duck: WGA-amplified duck genomic DNA as positive control; [2]Duck: duck genomic DNA as positive control; [3]No. Pos: number of hybrids positive for the assay (out of 90 hybrids tested); [4] Nb. Controls: total number of controls which are positive over the number of controls tested; [5]Mean pos: mean number of positive hybrids observed over the whole panel; NA: not applicable.

ADDITIONAL FIGURES AND TABLES

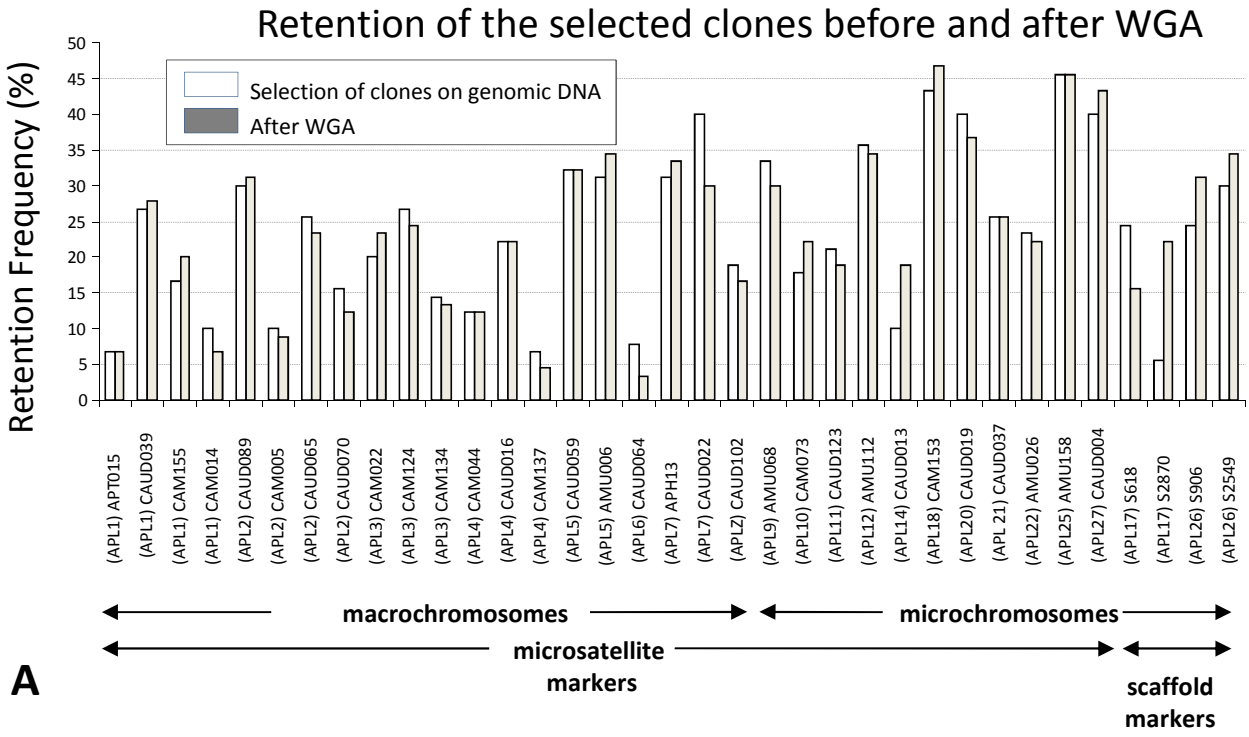
Additional File 1 Figure S1.pdf. Checking the 13 largest scaffolds by RH mapping. Each thick horizontal line represents a scaffold; arrows point to the names of the markers which were genotyped on the duck RH panel. The approximate position of the markers is shown as well as the scaffold lengths. Markers in the same color and contained within the same box are linked by RH mapping. For the 12 first scaffolds shown, the RH mapping data confirm the scaffold assembly. The last one, scaffold504, was the only one which was detected to be discontinuous, as marker sca504F is not linked by RH mapping to the five other markers sca504A, sca504B, sca504C, sca504D and sca504E. Comparative analysis with chicken shows that the portion of the scaffold containing sca504F aligns to GGA2, whereas the rest aligns to GGA1.

Additional File 2 Figure S2.pdf. Testing duck scaffolds aligning to two chicken chromosomes. Duck scaffolds are represented together with the portions of chicken chromosomes to which they show high sequence similarity in the Narcisse database. The approximate position of the markers on the scaffolds and on the chicken genome is shown as well as the scaffold lengths To test if the synteny breakpoints are due to an evolutionary chromosomal rearrangement or a problem in the assembly of scaffolds, a pair of markers was chosen close together on the scaffolds, but spanning the break points. Whenever markers are linked together by RH mapping, they are contained in the same box and are represented in the same colour.

Additional File 3 Table S1.xls. Genotyping results of 39 scaffold markers and 8 no hit ESTs for the three different methods. The panel contained 90 hybrids.

Additional File 4 Table S2.xls The 41 disrupted scaffolds which could be aligned on two different chicken chromosomes. Break1: the right-most coordinate of the alignment of the left part of the scaffold to one chicken chromosome. Break 2: the left-most coordinate of the

alignment of the left right part of the scaffold to another chicken chromosome. Chicken 1 and Chicken 2: chicken chromosomes to which the left and right parts of the duck scaffold align to respectively. Pair-end support: refers to the reliability of paired-end sequence data. **Additional File 5 Table S3.xls. Data on all markers genotyped in the study.** Primer pairs, PCR conditions, and accession numbers (where applicable) are given.


Retention of the selected clones before and after WGA
Retention Frequency (%)
Selection of clones on genomic DNA
After WGA
(APL1) APT015
(APL1) CAUD039
(APL1) CAM155
(APL1) CAM014
(APL2) CAUD089
(APL2) CAM005
(APL2) CAUD065
(APL2) CAUD070
(APL3) CAM022
(APL3) CAM124
(APL3) CAM134
(APL4) CAM044
(APL4) CAUD016
(APL4) CAM137
(APL5) CAUD059
(APL5) AMU006
(APL6) CAUD064
(APL7) APH13
(APL7) CAUD022
(APL2) CAUD102
(APL9) AMU068
(APL10) CAM073
(APL11) CAUD123
(APL12) AMU112
(APL14) CAUD013
(APL18) CAM153
(APL20) CAUD019
(APL 21) CAUD037
(APL22) AMU026
(APL25) AMU158
(APL27) CAUD004
(APL17) S618
(APL17) S2870
(APL26) S906
(APL26) S2549
macrochromosomes
microchromosomes
microsatellite markers
scaffold markers
A

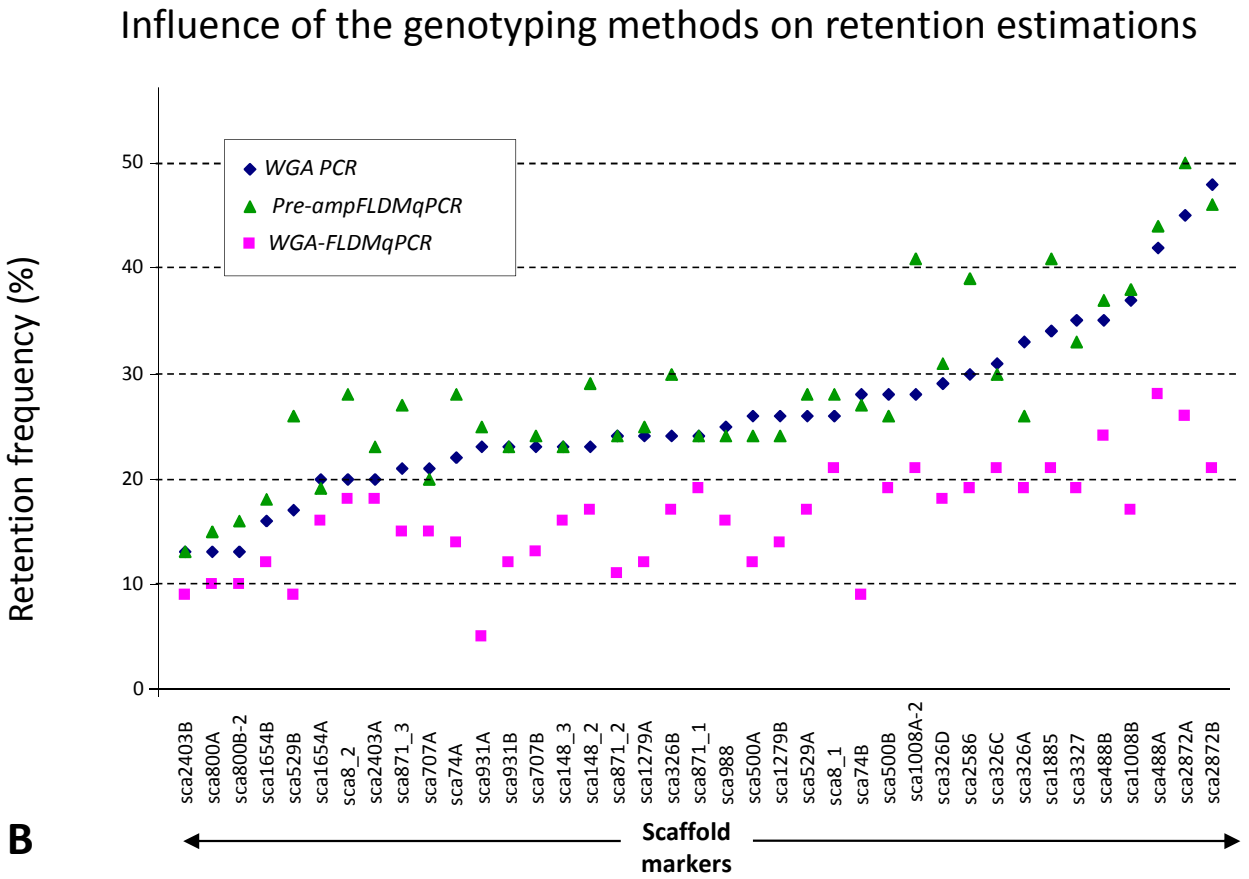
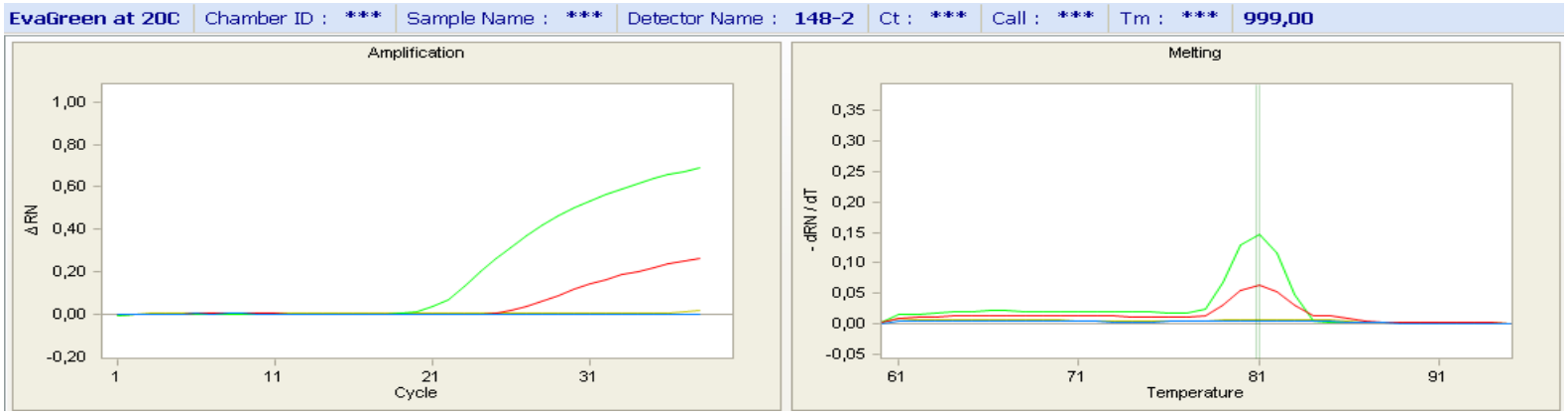

Influence of the genotyping methods on retention estimations
Retention frequency (%)
WGA PCR
Pre-ampFLDMqPCR
WGA-FLDMqPCR
sca2403B
sca800A
sca800B-2
sca1654B
sca529B
sca1654A
sca8_2
sca2403A
sca871_3
sca707A
sca74A
sca931A
sca931B
sca707B
sca148_3
sca148_2
sca871_2
sca1279A
sca326B
sca871_1
sca988
sca500A
sca1279B
sca529A
sca8_1
sca74B
sca500B
sca1008A-2
sca326D
sca2586
sca326C
sca326A
sca1885
sca3327
sca488B
sca1008B
sca488A
sca2872A
sca2872B
Scaffold markers
B

Figure 1

A


EvaGreen at 20C
Chamber ID : ***
Sample Name : ***
Detector Name : 148-2
Ct : ***
Call : ***
Tm : ***
999,00
Amplification
ΔRN
1,00
0,80
0,60
0,40
0,20
0,00
-0,20
1
11
21
31
Cycle
Melting
- dRN / dT
0,35
0,30
0,25
0,20
0,15
0,10
0,05
0,00
-0,05
61
71
81
91
Temperature

B

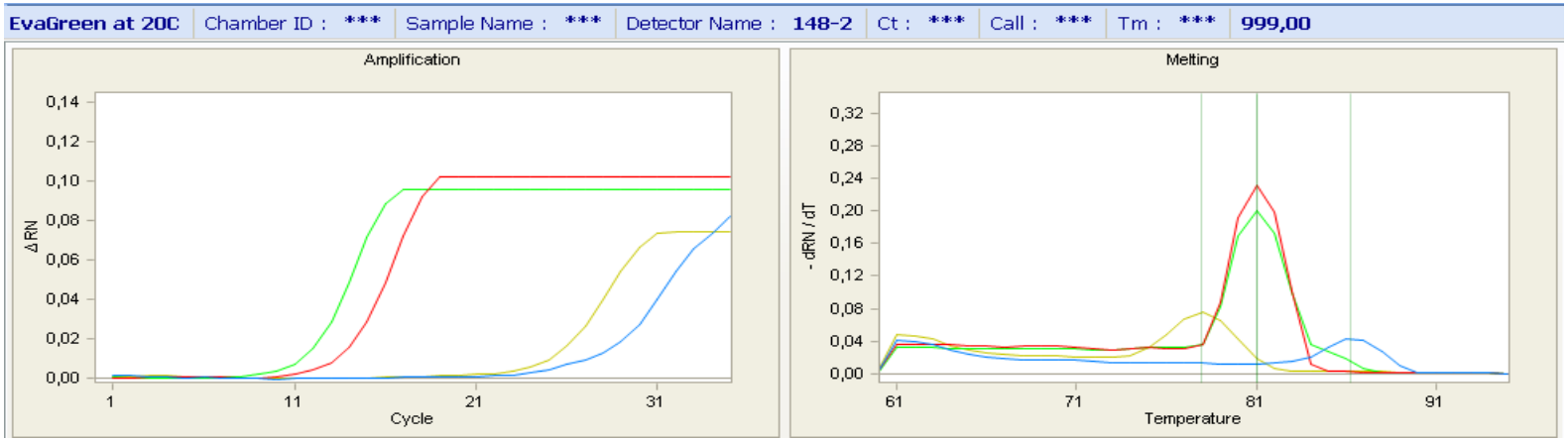

EvaGreen at 20C
Chamber ID : ***
Sample Name : ***
Detector Name : 148-2
Ct : ***
Call : ***
Tm : ***
999,00
Amplification
ΔRN
0,14
0,12
0,10
0,08
0,06
0,04
0,02
0,00
1
11
21
31
Cycle
Melting
- dRN / dT
0,32
0,28
0,24
0,20
0,16
0,12
0,08
0,04
0,00
61
71
81
91
Temperature

Figure 2

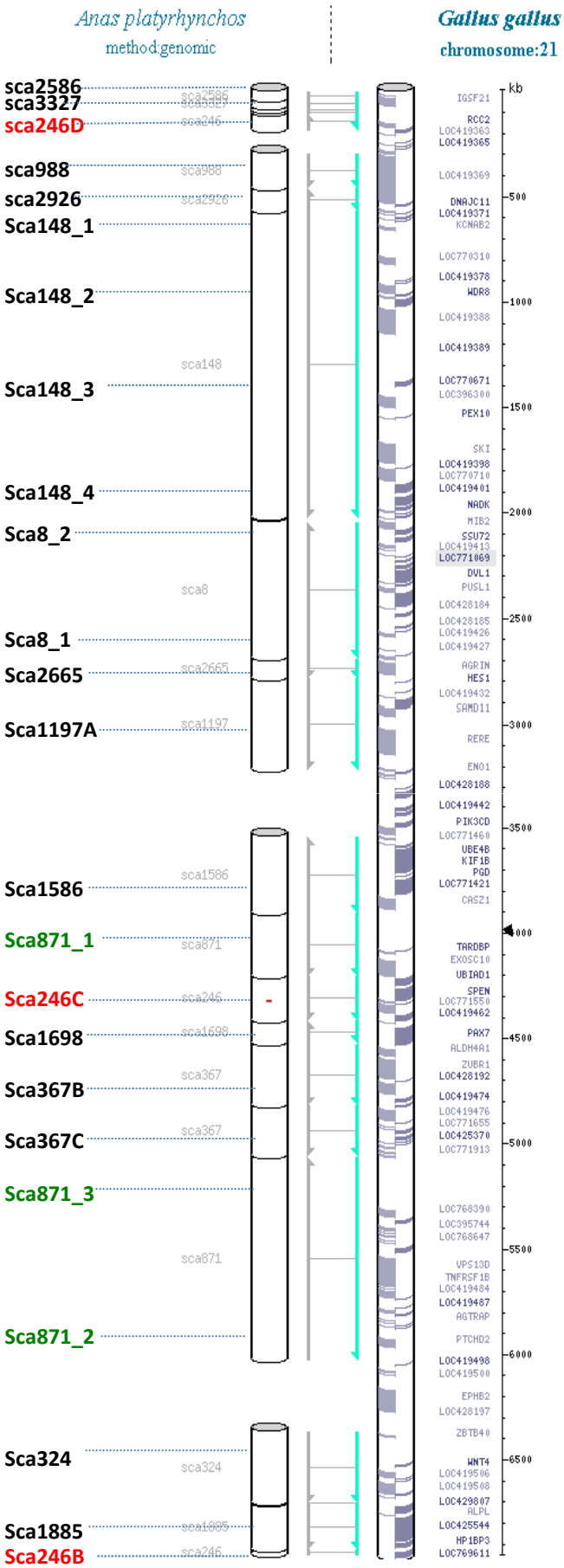

Anas platyrhynchos
method:genomic
Gallus gallus
chromosome:21
kb
sca2586
sca3327
sca246D
sca988
sca2926
Sca148_1
Sca148_2
sca148
Sca148_3
Sca148_4
Sca8_2
sca8
Sca8_1
Sca2665
Sca1197A
sca1586
Sca1586
Sca871_1
Sca246C
Sca1698
Sca367B
Sca367C
Sca871_3
sca871
Sca871_2
Sca324
sca324
Sca1885
Sca246B
IGSF21
RCC2
LOC419363
LOC419365
LOC419369
DNAJC11
LOC419371
KCNAB2
LOC770310
LOC419378
WDR8
LOC419388
LOC419389
LOC770671
LOC396300
PEX10
SKI
LOC419398
LOC770710
LOC419401
NADK
MIB2
SSU72
LOC419413
LOC771069
DVL1
PUSL1
LOC428184
LOC428185
LOC419426
LOC419427
AGRIN
HES1
LOC419432
SAMD11
RERE
ENO1
LOC428188
LOC419442
PIK3CD
LOC771460
UBE4B
KIF1B
PGD
LOC771421
CAS21
TARDBP
EXOSC10
UBIAD1
SPEN
LOC771550
LOC419462
PAX7
ALDH4A1
ZUBR1
LOC428192
LOC419474
LOC419476
LOC771655
LOC425370
LOC771913
LOC768390
LOC395744
LOC768647
VPS13D
TNFRSF1B
LOC419484
LOC419487
AGTRAP
PTCHD2
LOC419498
LOC419500
EPHB2
LOC428197
ZBTB40
WNT4
LOC419506
LOC419508
LOC429807
ALPL
LOC425544
HP1BP3
LOC769611
500
1000
1500
2000
2500
3000
3500
4000
4500
5000
5500
6000
6500

Figure 3

Figure 4

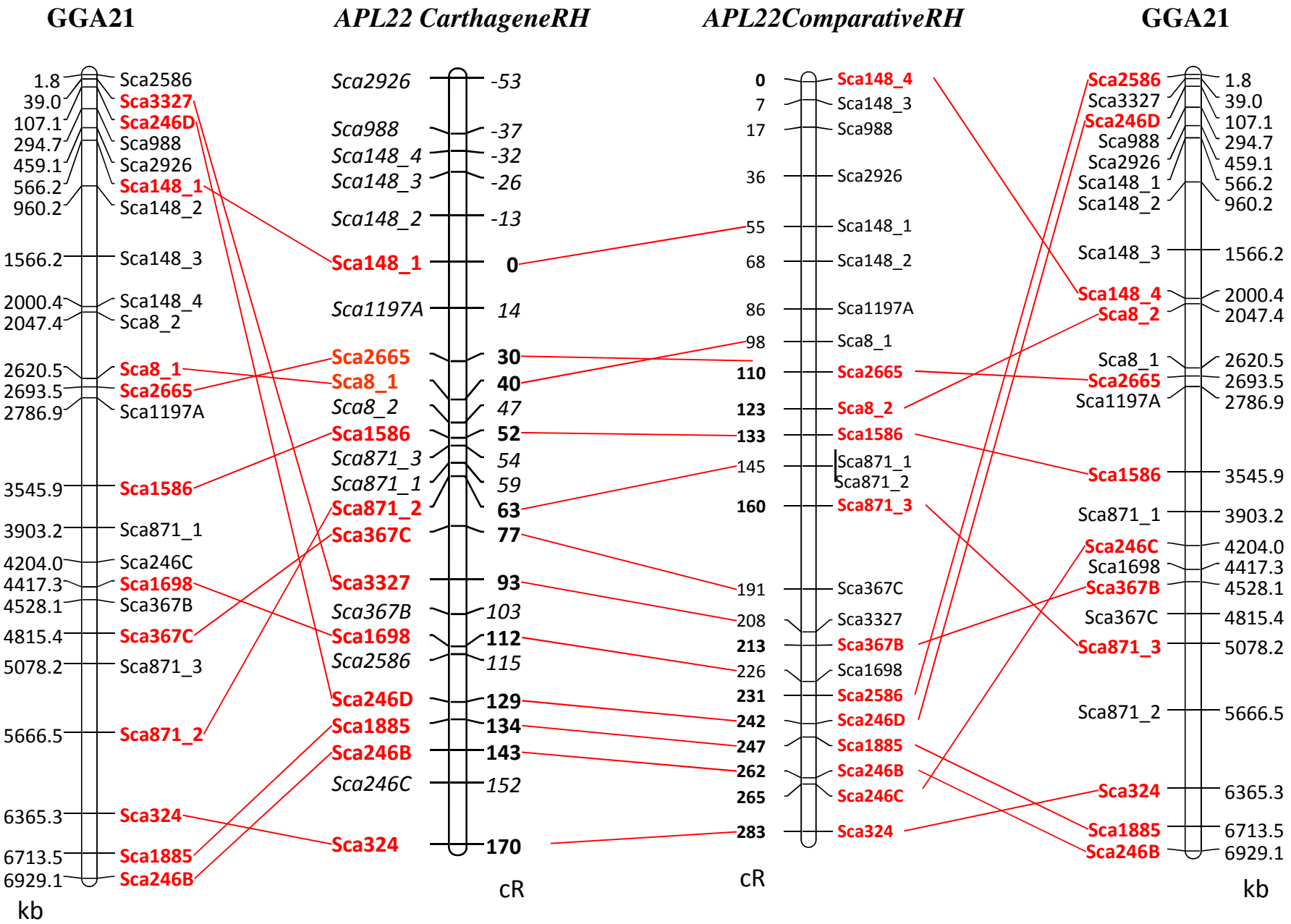

GGA21
APL22 CarthageneRH
APL22ComparativeRH
GGA21
1.8 Sca2586
39.0 Sca3327
107.1 Sca246D
294.7 Sca988
459.1 Sca2926
566.2 Sca148_1
960.2 Sca148_2
1566.2 Sca148_3
2000.4 Sca148_4
2047.4 Sca8_2
2620.5 Sca8_1
2693.5 Sca2665
2786.9 Sca1197A
3545.9 Sca1586
3903.2 Sca871_1
4204.0 Sca246C
4417.3 Sca1698
4528.1 Sca367B
4815.4 Sca367C
5078.2 Sca871_3
5666.5 Sca871_2
6365.3 Sca324
6713.5 Sca1885
6929.1 Sca246B
kb
Sca2926 -53
Sca988 -37
Sca148_4 -32
Sca148_3 -26
Sca148_2 -13
Sca148_1 0
Sca1197A 14
Sca2665 30
Sca8_1 40
Sca8_2 47
Sca1586 52
Sca871_3 54
Sca871_1 59
Sca871_2 63
Sca367C 77
Sca3327 93
Sca367B 103
Sca1698 112
Sca2586 115
Sca246D 129
Sca1885 134
Sca246B 143
Sca246C 152
Sca324 170
cR
0 Sca148_4
7 Sca148_3
17 Sca988
36 Sca2926
55 Sca148_1
68 Sca148_2
86 Sca1197A
98 Sca8_1
110 Sca2665
123 Sca8_2
133 Sca1586
145 Sca871_1
Sca871_2
160 Sca871_3
191 Sca367C
208 Sca3327
213 Sca367B
226 Sca1698
231 Sca2586
242 Sca246D
247 Sca1885
262 Sca246B
265 Sca246C
283 Sca324
cR
Sca2586 1.8
Sca3327 39.0
Sca246D 107.1
Sca988 294.7
Sca2926 459.1
Sca148_1 566.2
Sca148_2 960.2
Sca148_3 1566.2
Sca148_4 2000.4
Sca8_2 2047.4
Sca8_1 2620.5
Sca2665 2693.5
Sca1197A 2786.9
Sca1586 3545.9
Sca871_1 3903.2
Sca246C 4204.0
Sca1698 4417.3
Sca367B 4528.1
Sca367C 4815.4
Sca871_3 5078.2
Sca871_2 5666.5
Sca324 6365.3
Sca1885 6713.5
Sca246B 6929.1
kb

Figure 5

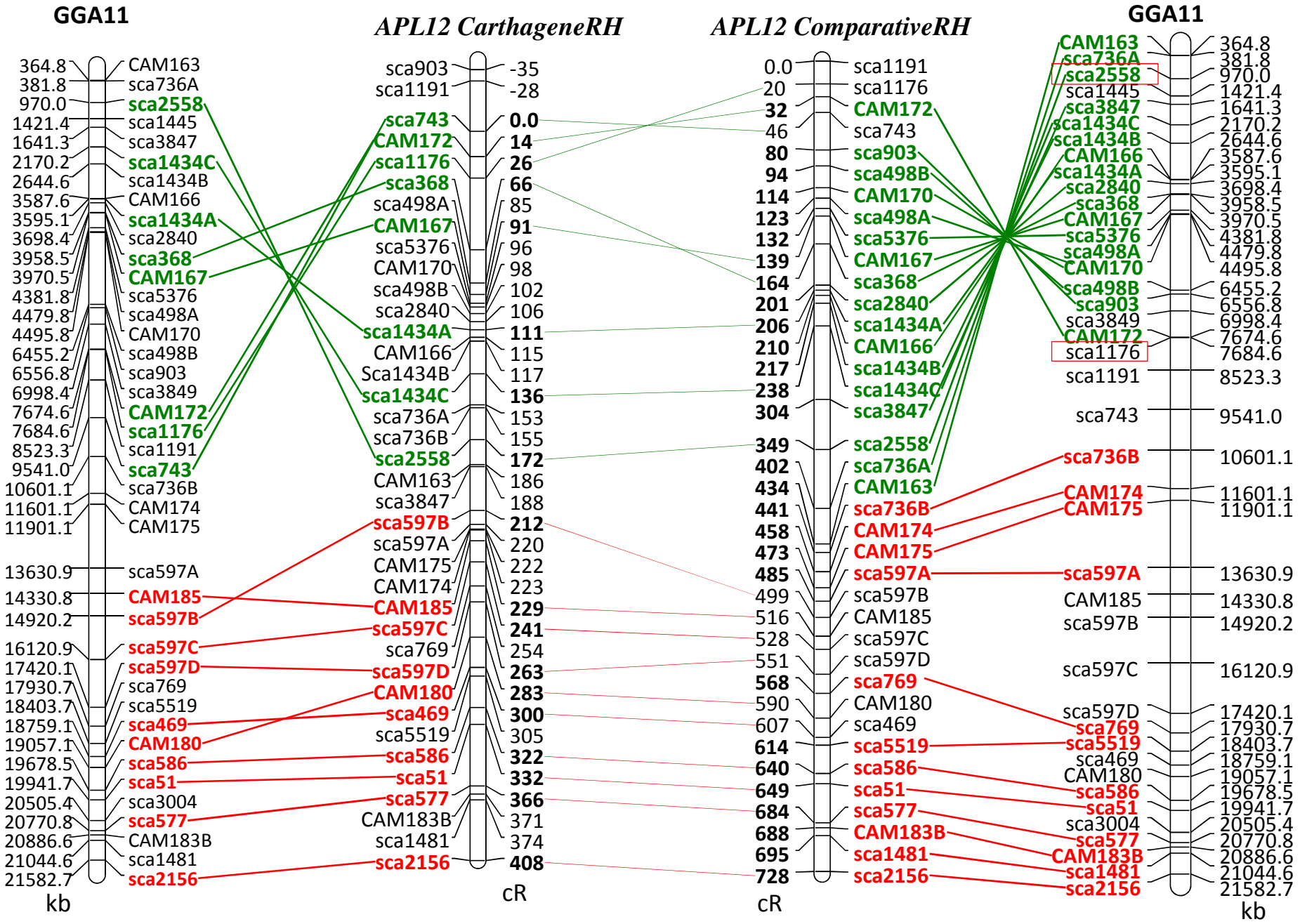

GGA11
364.8 CAM163
381.8 sca736A
970.0 sca2558
1421.4 sca1445
1641.3 sca3847
2170.2 sca1434C
2644.6 sca1434B
3587.6 CAM166
3595.1 sca1434A
3698.4 sca2840
3958.5 sca368
3970.5 CAM167
4381.8 sca5376
4479.8 sca498A
4495.8 CAM170
6455.2 sca498B
6556.8 sca903
6998.4 sca3849
7674.6 CAM172
7684.6 sca1176
8523.3 sca1191
9541.0 sca743
10601.1 sca736B
11601.1 CAM174
11901.1 CAM175
13630.9 sca597A
14330.8 CAM185
14920.2 sca597B
16120.9 sca597C
17420.1 sca597D
17930.7 sca769
18403.7 sca5519
18759.1 sca469
19057.1 CAM180
19678.5 sca586
19941.7 sca51
20505.4 sca3004
20770.8 sca577
20886.6 CAM183B
21044.6 sca1481
21582.7 sca2156
kb
APL12 CarthageneRH
sca903 -35
sca1191 -28
sca743 0.0
CAM172 14
sca1176 26
sca368 66
sca498A 85
CAM167 91
sca5376 96
CAM170 98
sca498B 102
sca2840 106
sca1434A 111
CAM166 115
Sca1434B 117
sca1434C 136
sca736A 153
sca736B 155
sca2558 172
CAM163 186
sca3847 188
sca597B 212
sca597A 220
CAM175 222
CAM174 223
CAM185 229
sca597C 241
sca769 254
sca597D 263
CAM180 283
sca469 300
sca5519 305
sca586 322
sca51 332
sca577 366
CAM183B 371
sca1481 374
sca2156 408
cR
APL12 ComparativeRH
0.0 sca1191
20 sca1176
32 CAM172
46 sca743
80 sca903
94 sca498B
114 CAM170
123 sca498A
132 sca5376
139 CAM167
164 sca368
201 sca2840
206 sca1434A
210 CAM166
217 sca1434B
238 sca1434C
304 sca3847
349 sca2558
402 sca736A
434 CAM163
441 sca736B
458 CAM174
473 CAM175
485 sca597A
499 sca597B
516 CAM185
528 sca597C
551 sca597D
568 sca769
590 CAM180
607 sca469
614 sca5519
640 sca586
649 sca51
684 sca577
688 CAM183B
695 sca1481
728 sca2156
cR
GGA11
CAM163 364.8
sca736A 381.8
sca2558 970.0
sca1445 1421.4
sca3847 1641.3
sca1434C 2170.2
sca1434B 2644.6
CAM166 3587.6
sca1434A 3595.1
sca2840 3698.4
sca368 3958.5
CAM167 3970.5
sca5376 4381.8
sca498A 4479.8
CAM170 4495.8
sca498B 6455.2
sca903 6556.8
sca3849 6998.4
CAM172 7674.6
sca1176 7684.6
sca1191 8523.3
sca743 9541.0
sca736B 10601.1
CAM174 11601.1
CAM175 11901.1
sca597A 13630.9
CAM185 14330.8
sca597B 14920.2
sca597C 16120.9
sca597D 17420.1
sca769 17930.7
sca5519 18403.7
sca469 18759.1
CAM180 19057.1
sca586 19678.5
sca51 19941.7
sca3004 20505.4
sca577 20770.8
CAM183B 20886.6
sca1481 21044.6
sca2156 21582.7
kb

Figure 6

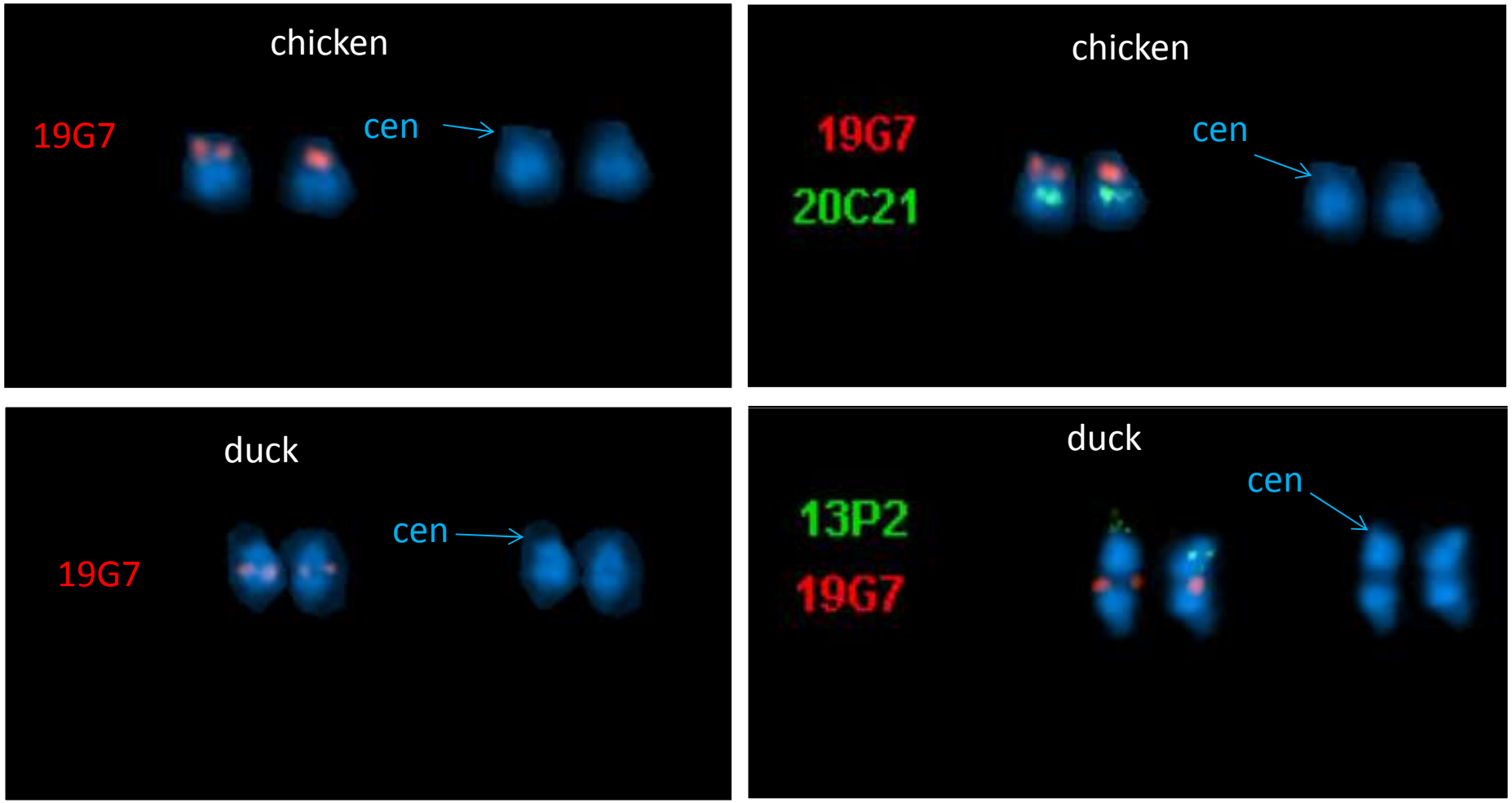
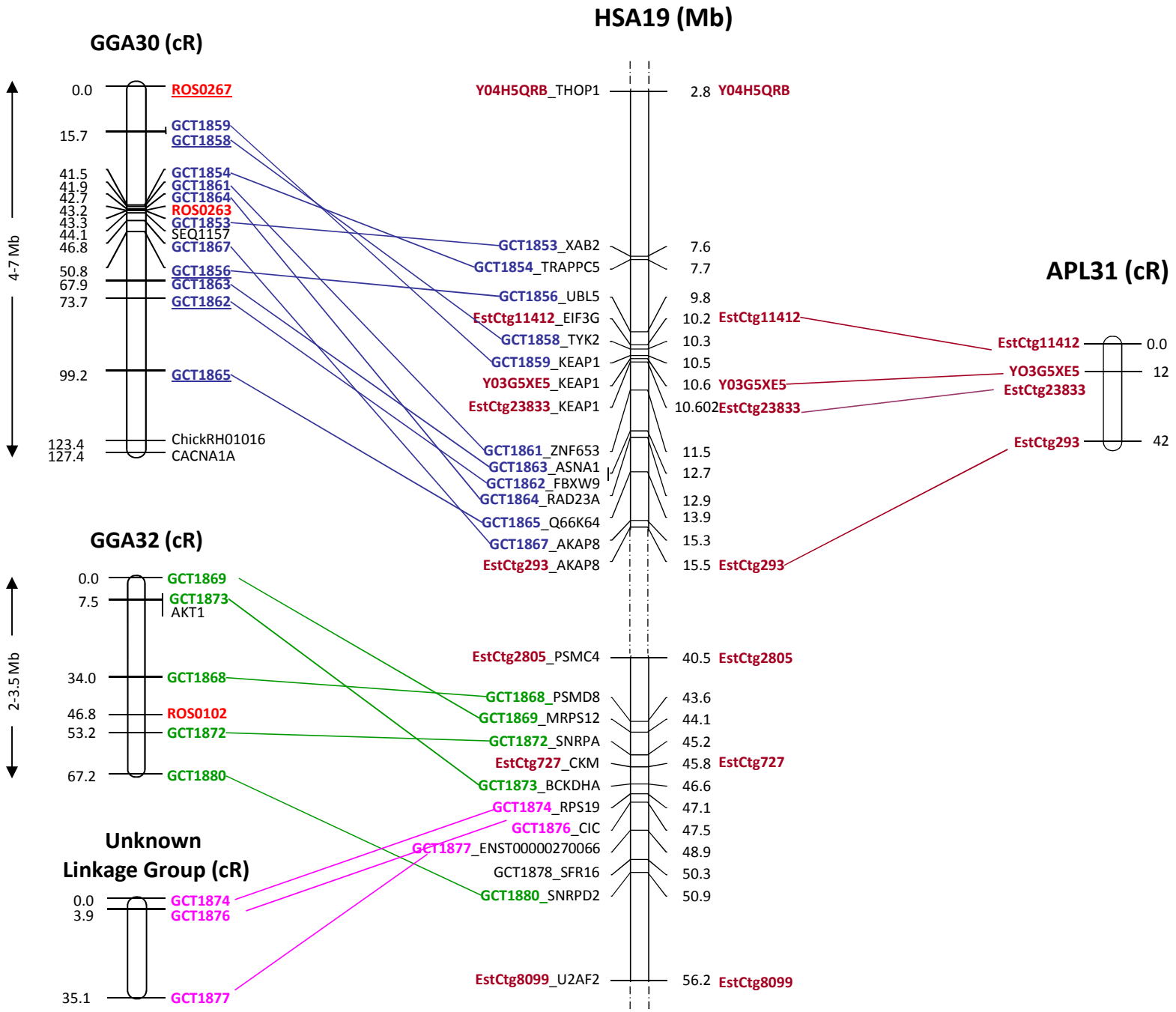

chicken
19G7
cen
chicken
19G7
20C21
cen
duck
19G7
cen
duck
13P2
19G7
cen

Figure 7


HSA19 (Mb)
GGA30 (cR)
4-7 Mb
0.0 ROS0267
15.7 GCT1859 GCT1858
41.5 GCT1854
41.9 GCT1861
42.7 GCT1864
43.2 ROS0263
43.3 GCT1853
44.1 SEQ1157
46.8 GCT1867
50.8 GCT1856
67.9 GCT1863
73.7 GCT1862
99.2 GCT1865
123.4 ChickRH01016
127.4 CACNA1A
GGA32 (cR)
2-3.5 Mb
0.0 GCT1869
7.5 GCT1873 AKT1
34.0 GCT1868
46.8 ROS0102
53.2 GCT1872
67.2 GCT1880
Unknown Linkage Group (cR)
0.0 GCT1874
3.9 GCT1876
35.1 GCT1877
Y04H5QRB_THOP1 2.8 Y04H5QRB
GCT1853_XAB2 7.6
GCT1854_TRAPPC5 7.7
GCT1856_UBL5 9.8
EstCtg11412_EIF3G 10.2 EstCtg11412
GCT1858_TYK2 10.3
GCT1859_KEAP1 10.5
Y03G5XE5_KEAP1 10.6 Y03G5XE5
EstCtg23833_KEAP1 10.602 EstCtg23833
GCT1861_ZNF653 11.5
GCT1863_ASNA1 12.7
GCT1862_FBXW9
GCT1864_RAD23A 12.9
GCT1865_Q66K64 13.9
GCT1867_AKAP8 15.3
EstCtg293_AKAP8 15.5 EstCtg293
EstCtg2805_PSMC4 40.5 EstCtg2805
GCT1868_PSMD8 43.6
GCT1869_MRPS12 44.1
GCT1872_SNRPA 45.2
EstCtg727_CKM 45.8 EstCtg727
GCT1873_BCKDHA 46.6
GCT1874_RPS19 47.1
GCT1876_CIC 47.5
GCT1877_ENST00000270066 48.9
GCT1878_SFR16 50.3
GCT1880_SNRPD2 50.9
EstCtg8099_U2AF2 56.2 EstCtg8099
APL31 (cR)
EstCtg11412 0.0
YO3G5XE5 12
EstCtg23833
EstCtg293 42

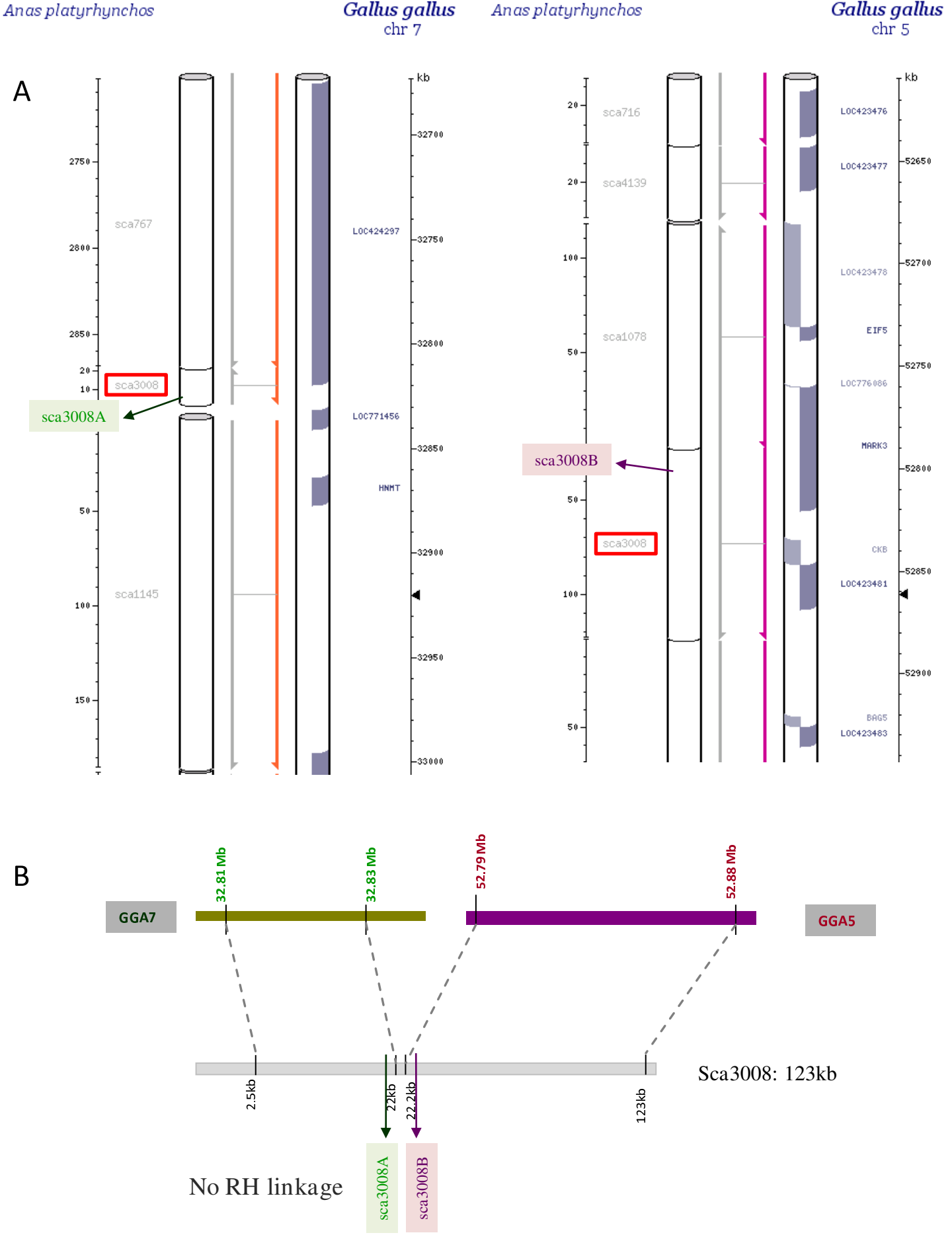

Anas platyrhynchos
Gallus gallus
chr 7
Anas platyrhynchos
Gallus gallus
chr 5
A
kb
sca767
sca3008
sca3008A
sca1145
LOC424297
LOC771456
HNMT
sca716
sca4139
sca1078
sca3008B
sca3008
LOC423476
LOC423477
LOC423478
EIF5
LOC776086
MARK3
CKB
LOC423481
BAG5
LOC423483
B
GGA7
32.81 Mb
32.83 Mb
52.79 Mb
52.88 Mb
GGA5
2.5kb
22kb
22.2kb
123kb
Sca3008: 123kb
No RH linkage
sca3008A
sca3008B

Figure 8

Additional files provided with this submission:

Additional file 1: Additional File 1 Figure S1.pdf, 58K
http://www.biomedcentral.com/imedia/1549460110751635/supp1.pdf
Additional file 2: Additional File 2 Figure S2.pdf, 375K
http://www.biomedcentral.com/imedia/4756895327516351/supp2.pdf
Additional file 3: Additional File 3 Table S1.xls, 37K
http://www.biomedcentral.com/imedia/1455493958661726/supp3.xls
Additional file 4: Additional File 4 Table S2.xls, 31K
http://www.biomedcentral.com/imedia/2918098666172693/supp4.xls
Additional file 5: Additional File 5 Table S3.xls, 124K
http://www.biomedcentral.com/imedia/1932062592661726/supp5.xls

Discussion

The two RH maps which were built proved the feasibility of using the panel as an aid for the duck genome assembly. Although the panel amplified by WGA proved inefficient for building maps for the smallest microchromosomes, it can still be used to construct the maps for other chromosomes by traditional genotyping.

By using sequence similarity, 1787 duck scaffolds were aligned to chicken chromosomes and could be visualized in the Narcisse database viewer (http://narcisse.toulouse.inra.fr/pre-narcisse/duck/cgi-bin/narcisse.cgi). However, many scaffolds whose length sum up to more than 51Mb could not be anchored by this comparative approach, either due to sequence divergence or to their very small size (Figure III-1). The information provided by Narcisse is the alignment of duck sequence on chicken chromosomes, meaning that the position and orientation of the scaffolds in the duck genome may be different. Whole genome comparison of zebra finch and chicken, whose divergence time is estimated to be around 100 MYA, have shown that there are extensive intrachromosomal rearrangements between two birds (Skinner and Griffin 2011; Warren et al. 2010). The divergence time between chicken and duck is estimated to be approximately 80 MYA (van Tuinen and Hedges 2001), but nevertheless, we estimate that there are more small scale intrachromosomal rearrangements yet to be identified, especially on medium size chromosomes and on microchromosomes. Therefore, whole genome RH maps will be an invaluable addition for the improvement of the current assembly and to facilitate other genetic approaches. Moreover, mapping scaffolds with the duck RH panel will allow the mapping of scaffolds that were too divergent to chicken in sequence composition to be included in the predictive Narcisse maps..

Many attempts have been made to sequence the smallest chicken microchromosomes using all available sequencing techniques (Ye et al. 2011) and despite this, their sequence is still absent from the current assembly of chicken. It has been suggested in many occasions, that high GC content and a high content in repetitive sequences of microchromosomes

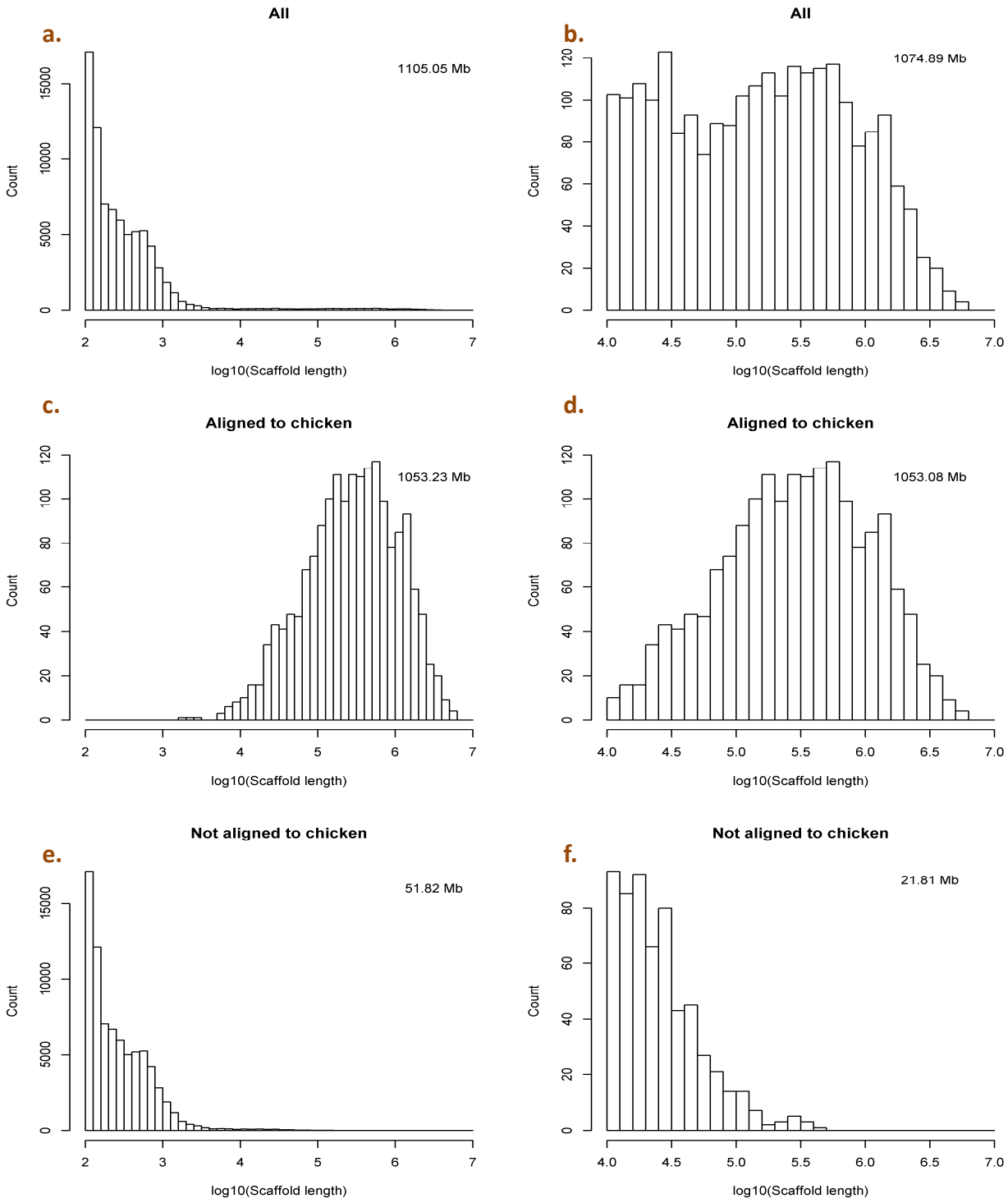


Figure III-1: Size distribution of duck scaffolds. Left: all 78 487 scaffolds; right: scaffolds larger than 10 kb. For each column: top: all scaffolds; middle: scaffolds aligned to chicken; bottom: scaffolds that do not align to chicken. Top left of each histogram: sum of scaffold length.

(McQueen et al. 1996; McQueen et al. 1998; Nanda and Schmid 1994) could cause difficulties in cloning and sequencing. However, these smallest microchromosomes are gene-rich and some EST data is available, that can be used as a start point for RH mapping. However, due to the fact that the exons of protein-coding genes are amongst the best conserved regions during evolution, cross-species amplification can happen and many markers are discarded due to cross-amplification of the hamster DNA in the hybrids when using traditional genotyping with PCR and agarose gel electrophoresis. This can be overcome by Fluidigm qPCR, as a Tm measurement of the PCR product is performed, allowing the distinction between specific and non-specific amplification. Available duck EST data is now available, but not at the same level as for chicken and despite the chicken RH panel was not amplified by WGA, large quantities of DNA were prepared. Thus, it would be wiser to continue the effort of characterization of the smallest microchromosomes in chicken, based on a much deeper sequencing and mapping of its genome. The Fluidigm qPCR genotyping of the chicken RH panel could help in this effort. Soon after the maps and sequence of the smallest microchromosomes in chicken are known, the smallest michromosomes in duck and other birds will possibly be inferred with the help of comparative maps and using the fact that synteny conservation is high.

Chapter IV.

Genotyping by Sequencing: whole genome RH maps

Introduction

We have reported a duck whole genome RH panel in the previous chapter and assessed three different genotyping techniques. Fluidigm qPCR showed a much higher throughput genotyping than traditional genotyping. To develop whole genome RH maps, according to our previous experience in mapping with the chicken RH panel, the minimum marker density should be about every 500kb, thus we need to map over 2,000 markers.

Designing 2,000 markers is very laborious, very expensive and time consuming, apart from marker design, the cost of Fluidigm microfluidic chip cannot be neglected. Although cheaper than traditional genotyping, the final overall cost of making whole genome RH maps would be very high: in the order of 40,000 € for 2,000 markers. Meanwhile, in the field of sequencing, constant improvements in technology have allowed regular decrease in cost and increase in throughput. At the time of writing moment, the cost for sequencing a large genome has dramatically decreased. Having in hand a preliminary assembly of the Duck genome and the duck RH panel mentioned above, we propose a new approach for RH mapping which consists in sequencing the RH panel as an alternative to genotyping and using duck scaffolds as markers to build maps and improve the duck genome assembly.

At first, our biggest concern was that the sequencing output would contain a majority of hamster reads which would be useless, with only approximately 3% of reads coming from duck (with 20% retention rate, a hybrid clone contains on average 200Mb duck genome compared to the 6Gb of the diploid genome of the recepient genome) . In a first approximation, if we set criteria of having a minimum of 4 independent reads to attest the presence of a scaffold in a hybrid, the sequencing of two million reads for each hybrid led to a risk of 0.0003 of missing a scaffold whose length is at least 50kb. Considering this risk is acceptable, a total of 18Gb sequences for 90 hybrids should allow mapping a majority of the scaffolds larger than 50kb, thus covering 95.7% of the assembly. This stimulation provides a

first evidence that a survey sequencing of the RH panel, combined with RH mapping would extract maximal genome information for duck.

We have sequenced 100 hybrids at a mean coverage of 0.3X and proved that survey sequencing at this depth allowed construction of whole genome RH maps. Two thousand and twenty seven scaffolds were placed on 27 chromosomes and thereafter provided opportunity to compare with chicken genome.

Article in preparation

Here is only a preliminary version of the manuscript for readers.

Sequencing Radiation Hybrids for Improving Genome Assembly: Example of the duck genome

Introduction

The ultimate goal of genetics is to link each phenotype to genotype which resides in a genome. Therefore, a complete genome is an invaluable repertoire for biomedicine approach, evolutionary study and animal/plant breeding. The length of the sequencing reads (100 ~1000bases) produced by any state of the art sequencing technology is by several orders of magnitude smaller than the genome size. Thus any genome sequencing project involves a strategy to assemble the sequencing reads into complete genome. The sequence assembly process is a stepwise process in which the sequence reads are first organized into contiguous sequences (contigs) and subsequently in larger structures called scaffolds, the latter being finally ordered and oriented on the different chromosomes using external mapping information.

The first genome projects, most notably the human genome project (Lander et al. 2001), followed a “map first, sequence second” strategy, also known as clone-by-clone (CBC) method (Green 2001), where a physical map constructed beforehand is used to select the clones to be sequenced and hence to organize the contigs along the chromosomes. The alternative whole genome approach, the whole genome shotgun (WGS) method, proceeds more directly by the assembly of sequence reads generated in a random, genome wide fashion. While this approach bypasses the labor intensive construction of a clone-based physical map, it doesn’t dismiss however the need for genome wide maps in order to organize the resulting assembly contigs along the chromosomes. Since the year 2007 the massive parallel sequencing, or so called next generation sequencing (NGS), have revolutionized genomics by its unprecedented speed, throughput and ultra low cost. All NGS technologies sequence a genome routinely by means of WGS. Hitherto many non-model species have been sequenced using one or several NGS technologies, to name a few, the giant panda (Li et al. 2010), duck (Huang and consortium 2012) and yak (Qiu et al. 2012). Due to the limited mapping ressources available for most non-model species, the genome assemblies have been or will be published as a collection of scaffolds which are not organized on chromosomes. Whatever the sequencing strategy, the top down clone-by-clone method (Green 2001) or the more widespread Whole Genome Shotgun (WGS) method , long range intermediate genome maps

are needed to organize the contigs and scaffolds along the chromosomes. Besides the BAC FPC maps that are central to the clone-by-clone approach and the widespread genetic maps, the RH maps are also commonly used in this physical mapping process of assembling contigs into chromosomes. RH maps have been extensively used to assist the assembly of dog (Lindblad-Toh et al. 2005), cat (Pontius et al. 2007) and bovine (Elsik et al. 2009) to name only a few.

Every genome sequence needs a good map" (Lewin et al. 2009). Harris Lewin and his colleagues emphasize the importance of having physical maps with good resolution for optimizing utilization of genome sequences generated by WGS approaches. Indeed, Many species have already been sequenced by WGS sequencing, along with high density intermediate maps; their genome sequences are available in the form of chromosomes (Dalloul et al. 2010; Gibbs et al. 2004; ICGC. 2004; Lindblad-Toh et al. 2005; Warren et al. 2010; Waterston et al. 2002). For instance, dog was sequenced by a WGA approach, from which an improved assembly CamFam2.0 had a N50 scaffold of 45.0Mb in length whereas the total assembled size was 2.385 Gb (Lindblad-Toh et al. 2005). The high density integrated RH/FISH maps provided invaluable data helping to anchor the assembly to the canine chromosomes (Breen et al. 2004; Breen et al. 2001) and allowing 97% of the assembly to be ordered and orientated. Such an improved genome assembly, approaching the reality of the chromosome ordering, is indispensable for comparative genomics.

The Duck genome has been recently sequenced by NGS with Illumina *GAII* sequencing machines. A total of 78,487 scaffolds have been assembled, with a scaffold N50 of 1.2Mb and the largest being 5.9Mb in length (Huang *et al, in prep*). The current duck mapping resources are quite limited (Huang et al. 2006). To this end, we propose a high throughput RH mapping method to order and orientate the NGS assembly, using duck as an example and the RH panel recently developed in our laboratory (Rao et al. 2012). This duck panel has an average retention of 23% and already showed its power in assisting NGS genome assembly (Rao et al. 2012). The state of the art genotyping methods for a RH panel are PCR-based or chip-based. Currently, we neither have sufficient markers developed for PCR-based genotyping nor a ready-to-use chip. Therefore we decided to sequence the RH panel to order scaffolds and accomplish the chromosome assignment in duck. To better understand the method, we first describe rationale in result section.

Results

Rationale

In broad outline, a radiation hybrid panel is constructed by randomly fragmenting the genome by irradiation and rescuing a subset of the resulting fragments in a recipient cell. The proportion of fragmented genome rescued is called the retention fraction and the breakage frequency between two markers is simply the proportion of hybrid cell lines in which a breakage occurred between the two markers The chromosomal breakage induced by the radiation plays here a similar role as recombination in genetic mapping, the probability that two linked genes are included within a single fragment, and therefore their co-retention probability, decreases with the distance between them. Key to success of RH mapping is the ability to determine correctly the retention pattern of markers – e.g. the presence/absence status in all the clones. In the absence of a large collection of markers in duck, having in hand a RH panel and the NGS assembly described above, we propose to sequence the 100 hybrid clones of the duck RH panel enabling to genotype directly the assembly scaffolds on the panel. The rationale is as follow: the presence/absence of a scaffold in a particular sequenced hybrid is attested by the presence/absence of reads mapping specifically to this scaffold. We describe in the following section, the different steps of this mapping by radiation hybrids sequencing procedure: from the raw sequence data to the retention pattern for the scaffold-markers to the construction of the maps.

Primary data

We have sequenced 100 RH hybrids and a total of 179 Gb sequences were produced. Considering a mean retention of 20% of a haploid duck genome in each hybrid and haploid genome sizes of 1 and 3 Gb for duck and hamster respectively, a hybrid clone contains 6Gb of hamster and 200 Mb of duck DNA (3 % of duck DNA per hybrid). Thus the mean expected coverage of duck genome sequence per hybrid is 0.3 X. The amount of sequence produced for each clone is indicated in Supplementary Table1. There were clear biases in the results, as the quantity of sequence produced per hybrid could vary up to 8-fold, with the minimum amount for hybrid h215 having 2 million reads and the maximum for hybrid h225 having more than 17 million reads. The average percentage of the reads which can be uniquely aligned on the duck scaffolds is around 2.5 %, which is close to our expectation of 3 % when considering an

average retention of 20 % of the duck genome in the hybrids. Here we define read coverage as number of paired reads per 20kb which was an important parameter in genotype calling process in following analysis. In our initial estimations, considering 20% retention of haploid genome retained and without sequencing biases, we expected a mean value of 3 reads per 20kb for duck scaffold. We plotted the length of scaffolds having at least one pair of reads well and uniquely mapped within each hybrid and found that many scaffolds only had very few read pairs mapped (less than 1 read per 20kb). We also plotted the length of scaffolds having at least 1 read per 20kb, resulting in a significant loss of positive scaffold in all hybrids and dramatic losses in some hybrids (i.e. h100) as show in Supplementary Figure1A. To further investigate the reasons for which there were so many scaffolds with such extremely low read coverage, we visualized data with GenomeView (Abeel et al. 2012) and discovered that some scaffolds could have read pairs clustered in specific region.

In traditional RH genotyping by PCR, care was taken in marker design to avoid nonspecific amplification of the hamster genome, leading to false positive calling. Similarly, we filtered sequencing reads that could be mapped both on duck and hamster, resulting in an average of 1.6% sequencing reads left for analysis (Supplementary Table1). The successfully mapped read pairs for each hybrid varied between 16 521 to 408 453, almost differing by 25-fold. We plotted the scaffolds having at least 1 reads and having the read coverage of at least 1 read per 20kb after removal of potential hamster reads, and showed the data for three example hybrids in Supplementary Figure1B. In both datasets, we found that the in hybrid containing few reads mapped, i.e. h100, the proportion of scaffolds containing at least 1 read per 20kb was very low. Therefore, we observe that a lower read coverage leads to more ambiguities in the determination of presence or absence of smaller scaffolds. Due to this problem, 14 hybrids highlighted in grey in Supplementary Table1 were excluded from further analyses.

Read coverage varies within and among hybrids

To better understand the data, we plotted read coverage of the scaffolds having read coverages of at least one read mapped, at least 1 read per 20kb and at least 4 reads per 20kb before and after filtering potential hamster reads respectively. Three hybrids are shown as an example in Figure1, from which we could conclude that there was a great proportion of scaffolds with very low read coverage (less than 1 reads per 20kb) in all hybrids. For scaffolds having at least 1 read or 4 reads per 20kb, there is a clear bimodal distribution suggesting that

duck fragments are not uniformly distributed. Indeed, we estimated the read coverage distribution at similar sequencing depth using chicken sequencing data (F.Pitel personnel communication) and then compared with duck hybrid sequencing data, shown in Figure 2, from which a more dispersed distribution was observed in duck hybrid sequencing data. This suggests strongly that the hybrid cell lines are in fact a mixture of cells retaining different duck chromosome fragments, with some fragments retained in a majority and perhaps all cells, whereas others are in a minority of cells. This is consistent with data from human classical somatic hybrid cell lines, demonstrating that a single hybrid cell line is not homogenous but a mixture of different cells. As an example, table 2 shows the percentage of cells from a human chromosome assignment panel containing given chromosomes. (A.Vignal personnel communication) Supplementary Table 2. Figure1 shows that read coverage within a single hybrid varies significantly from more than 350 reads per 20kb down to less than 1 read per 20kb. As shown in Figure1, h100, that was eliminated from the subsequent analyses, contains a majority of scaffolds with low read coverage, and hybrids which were kept contain a considerable higher number of scaffolds with a high read coverage scaffolds. Thus in the final dataset of 86 hybrids, the analysis should generate a lower false calling rate. An example of read coverage along a scaffold is shown in Supplementary Figure2 for sca519. Read coverage can vary on the same scaffold within a hybrid like in h154; a clear breakage observed in h156 (Supplementary Figure 2).

When we compared the read coverage for the same scaffold among different hybrids, we observed variations between hybrids which are completely independent from the quantity of sequence obtained for the hybrids. For instance, although h102 is amongst the hybrids with the highest sequencing output (Supplementary Table 1) and the highest overall retention rate (Rao et al. 2012), it has a lower density of reads compared to other hybrids for sca109 (Supplementary Figure 3). This suggests that the proportion of cell containing the duck chromosome fragment corresponding to sca109 is lower in h102 than in the other hybrids. Sca109 is present as two independent fragments in hybrid h295.

Scaffold Segmentation

As shown in Supplementary Figure 2, breakages can be observed within scaffolds in the hybrids. To detect such breakpoints in the entire dataset, we used segmentation algorithms. We first used the Circular Binary Segmentation (CBS) algorithm, which was primarily

developed to detect copy number variations (Olshen et al. 2004), in order to segment scaffolds with a window size of 20kb. CBS is a modification of a binary segmentation algorithm which was based on a test to detect a single change (breakpoint) (Sen and Srivastava 1975). On the contrary, there is no limitation on the numbers of changes (breakpoints) that can be detected in CBS.

Nevertheless, CBS requires a defined sliding window size and thus the window containing a breakpoint can be assigned to the wrong side if the read coverage is significantly lower than neighboring window, which can result in increasing the read coverage on the absent side (Supplementary Figure4). In addition to breakpoint imprecision, CBS also fails to detect breakage in some rare cases (Supplementary Table3).

Segments Calling

To summarize the segmentation of the scaffolds in the hybrids and to format results, we only considered the segmentation in two ends of each scaffold, meaning that when a scaffold was segmented in more than two we only used the segmentation results from the two ends, to allow the orientation of the scaffolds in the maps. We used a threshold of < 0.5 read per 20kb for calling genotypes as absence of scaffold or of scaffold ends in hybrids; of > 1 reads per 20kb for presence scaffold or of scaffold ends in hybrids; values in between were called as unknown. Scaffolds having identical genotypes at both ends for all the 86 hybrids were considered as a single marker, as no breakage was observed. Contrarywise, if a scaffold had differents genotypes at both ends in at least one hybrid, it was treated as two markers, allowing for possible orientation on the RH maps. To eliminate bad quality markers, we selected only those having a retention higher than 5% in the panel and an unknown calling rate less than 15%. The final dataset is composed of 2690 markers from 2027 scaffolds covering 1055 Mb of the duck genome assembly.

Linkage analysis

We performed linkage analysis using the set of 2690 markers, from which 51 linkage groups were obtained using a LOD score threshold of 4.5. We superimposed these 51 linkages on chicken chromosomes, represented in Figure4. These results suggest a good agreement with the cytogenetic data, confirming that no interchromosomal rearrangement can be detected, except for GGA4 corresponding to APL4 and APL10.

The Maps

Traditional RH mapping constructs the maps on the sole information given by the RH vectors (de Givry et al. 2005). Although this method is suitable for genotyping a small number of markers, it becomes tedious when several hundreds or thousands are involved. Therefore we built the duck RH maps using a comparative mapping approach suitable for genome-wide marker ordering (Faraut et al. 2007), in which a genome phylogenetically close to the genome to be mapped is used as a reference to help in marker ordering. To check if the final result can be influenced by the reference genome, we built 3 successive sets of maps on our segmentation and calling results by using chicken, zebra finch or turkey as reference genomes. In traditional RH mapping, a framework map is a map whose marker ordering is 1000 times more reliable than any other ordering with the same set of markers, whereas in the comparative mapping approach, the map containing a set of marker with an invariant order is named a robust map.

In this study, we focused on three chromosomes: APL2, APL12 and APL22. The rationale for selection of these chromosomes is the following: APL2 was a large chromosome in which cytogenentic data showed some unelucidated rearrangements (Fillon et al. 2007; Skinner et al. 2009) and reported to have a well conserved synteny with turtle chromosome 2 (Graves, unpublished data) , APL12 and APL22 had already been genotyped with different techniques (Rao et al. 2012). These three chromosomes are in addition good representatives of typical avian macrochromosomes, minichromosomes and michrochromosomes.

APL22

We used the chicken genome as a reference to construct the RH map for this microchromosome (Figure 5). This map (APL22_GGA21) contains 22 markers corresponding to 16 duck scaffolds, in which 12 markers are placed on the 198.8cR long robust map. The average retention for APL22 is 31% and retentions of the region spanning from marker sca246_1 to sca324_0 were the highest, suggesting the centromere could be in this region (Benham et al. 1989; Goodfellow et al. 1990). This map suggested the existence of many complex rearrangements between chicken and duck.

APL12

For this minichromosome we used as reference the three genomes chicken, turkey and zebra finch genome. The map using chicken as reference (APL12_GGA11) is 358.9cR long and consists in 47 markers, 39 of which placed on the robust map, which is highly consistent with the previous map made by comparative mapping (APL12_FLDM) (shown in Figure 6). The map made by using zebra finch as reference (APL12_TGU11) contains 35 markers on the robust map and agrees highly with APL12_GGA11 (shown in Figure 7). This map is more than 400cR long and contains 45 markers. The map made by using turkey as reference (APL12_MGA13) contains 40 markers with 36 placed on the robust map, and is 353.8cR long (Figure 8).

The average retention is 38.4% and the highest retention is for the first marker (sca743_0), suggesting that the centromere could be close to this region. This is in agreement with the cytogenetic data, also showing that APL12 is a telocentric chromosome (Fillon personnal communication).

There is a major intrachromosomal rearrangement between APL12 and GGA11, which was confirmed by FISH experiment in a previous study, involving about 10Mb. Comparing APL12 with zebra finch, shows three rearrangements, including one translocation (Figure 7). When compared with APL12_GGA11, APL12_MGA13 has five scaffolds counting as 7 markers, highlighted in yellow in the figure, which were less conserved between duck and turkey and couldn't be located by sequence similarity on the turkey assembly and therefore couldn't be used as markers in the comparative mapping approach. Two inversions and one tranlocation were revealed by comparative mapping between duck and turkey, aas shown by arrows on the figure.

Comparing three RH maps made by using different references, there were some markers not in common. For instance, sca575 was placed on ChrUN (chromosome unknown) in chicken but on TGU11 in zebra finch. The lengths of three maps were quite similar after removing the marker sca575.

APL2

The map made with chicken as a reference (APL2_GGA2) contains 328 markers, of which 319 are placed on the robust map. This map is 1688,3cR long and the sum of the

scaffold lengths is about 158Mb, similar to the length of GGA2 (Figure 9). The map using the zebra finch genome as reference (APL2_TGU2) consists of 308 markers with 296 assigned on the robust map (Figure 10). The robust map is 1658.4 cR long and coveres 156.6 Mb of the duck genome which is similar to the length of TGU2. To use the turkey genome as a reference, we concatenated MGA6 and MGA3 as a virtual chromosome to construct a RH map for APL2 (APL2_MGA) as cytogenetic data suggested that fission of ancestral chromosome 2 gave rise to MGA3 and MGA6 (Dalloul et al. 2010; Zhang et al. 2012). APL2_MGA consists of 278 markers, 265 of with assigned to the robust map(Figure 11). The comparative map is 1568.4 cR long and covered 142.1 Mb of duck genome.

We also compared the robust maps for APL2 constructed with the 3 different reference genomes. There were 238 markers in common and all maps showed a high degree of consistence (Figure 9,10,and 11). We plotted the retention for all the markers on the RH map obtained from the chicken-duck comparative mapping in Figure 12, from which a clear centromere effect was observed. The average retention is 18.2% and we suggested that the centromere could be close to sca1153_0.

The RH map of APL2 suggests 7 tentative intrachromosomal rearrangement when compared to GGA2: 6 inversions and a large translocation. The largest inversion from sca1034_0 to sca74_1 on GGA2, spans about 11Mb. Interestingly, this rearrangement is supported by assembly scaffold as well, as this inversion led to two duck scaffolds (sca74 and sca1034) to be splitted when aligned to GGA2. The second largest inversion involves the chicken centromeric region: about 10Mb, between sca2872 and sca616_0. The translocation from sca713 to sca616_1, transposes approximately a 6 Mb fragment from the q arm to the p arm. Comparing with zebra finch, 4 putative inversions and 4 putative translocations are suggested.

All maps showed that two inversions are shared between the three comparative maps, specified by the orange box in the Figure9, 10 and 11, a fragment of approximately7Mb from sca258_0 to sca22_1 and another fragment of 6 Mb from sca5_0 to sca280_0. These seem to be duck-lineage specific inversions.

In addition, by integrating the previous cytogenetic data on the comparative maps between chicken and duck, the RH map confirmed the complex rearrangements on this chromomosome (data in Figure13)(Fillon et al. 2007; Skinner et al. 2009). Due to the fact that random selection of BAC clones in these studies, some were selected just outside the

rearrangement, such as the BAC corresponding to sca9452 and sca1034, or in the middle of inversion such as the BAC corresponding to sca1153, Current FISH data could not illustrate the complex intrachromosomal rearrangement explicitly. To this end, we selected 4 chicken Wageningen BAC clones to perform FISH experiments, which confirmed the translocation and the inversion from marker sca1034_1 to marker sca74_1 (Figure 14).

Disrupted Scaffolds

In the previous study, we demonstrated that we detected 41 scaffolds, also called disrupted scaffolds, that could be mapped to two different chicken chromosomes by Narcisse (Courcelle et al. 2008; Rao et al. 2012). Birds are were well known to have very well conserved karyotypes and syntenies, and that there were no interchromosomal rearrangements detected to date between chicken and duck except for GGA4 corresponding to APL4 and APL10 (Fillon et al. 2007; Skinner et al. 2009). In the previous study we only chose 19 scaffolds as a survey study to test the power of the newly made duck whole genome RH panel and to test these potentially misassembled scaffolds. All cases were proved to be misassembled scaffolds, except for sca649. From the genotype calling generated from sequencing hybrids, most of scaffolds larger than 20kb were kept and all the 41 scaffolds were called, among which sca180 and sca649 were proven to be correctly assembled and rest were confirmed as misassembled except for uncertainty of sca398 and sca802 (shown in Supplementary Figure5); all data on the potentially misassembly regions is summarized in Supplementary Table5. Both sca398 and sca802 were mapped to sexual chromosomes in chicken. CBS segmentation suggested both scaffolds were in the same linkage group whereas in the graphical representation of Seqmonk (http://www.bioinformatics.babraham.ac.uk/projects/seqmonk/) the breakages, if real, are not obvious, due to low read coverage (Supplementary Figure5)

Moreover, we also found one case (sca530, mapped to GGA3), in which the comparative data didn't suggest a misassembly, although one was revealed by linkage analysis, at the end of the scaffold (Supplementary Figure6).

Discussion

As shown in Figure1, each hybrid exhibited its own pattern in read coverage. But there were some characteristics shared in common: all hybrids contained a great proportion of

extremely low read coverage reads and the rescued scaffolds were contained at different levels in term of copy number (data not shown). We gathered the low read coverage scaffolds from different hybrids and tried to find some clues, but it seemed that there was no clear pattern for the low read coverage scaffoldss which meant that the low read coverage scaffoldss were not shared in all hybrids. For each hybrid, the distribution of read coverage was complex and prone to be overlap of two or more negative binomial distribution. The mechanism for this phenomenon is not yet known. We speculated that by nature that a hybrid cell line was a mixture and that random loss of fragments might be responsible for the complicated distribution of read coverage, and the contamination in sequencing could also be a reason especially for low read coverage scaffoldss. The nature of heterogeneity of a hybrid cell line was not only supported by human somatic hybrid cell (Supplementary Table2), but also evident in our results from the characterization of hybrids(FIGURE14). For instance, in h207, some cells contained only one synthetic microchromosome while some had nine synthetic microchromosomes which were composed of duck fragments. A question was thus raised whether this cell was not a single clone at the moment of the colony isolation or that the great variation in synthetic microchromosomes was as a result of fragment random loss. Both situations could perhaps exist in the hybrids; moreover, the hybrid cells were passaged no more than 4 generations which could lead to incomplete loss of the duck fragments, and might explain the low read coverage scaffolds in the data. It would be interesting to investigate some chicken radiation hybrids (Morisson et al. 2002) which were subject to large scale culture to understand better the phenomenon. The selective gene, HPRT, was always among scaffolds that had the highest read coverage. To assess the percentage of the cell containing a scaffold, it would be better to use the average read coverage of hamster genome instead of selective gene HPRT although the recipient cell line hamster $Wg3hCl_2$ was transformed. As it was reported that selective gene could amplify under selection pressure (Carroll et al. 1988; Carroll et al. 1987; Schimke 1984; Stark 1986), we did observe gene amplification of HPRT in one case in h304 so that the estimating proportion of cell containing a given scaffold could be imprecise. An intriguing question can be raised with regard to the telomeres in the hybrids, as shown in FIGURE14. Duck fragment are preferentially rescued by forming synthetic microchromosomes and no telomere effect was reported, then without protection of the telomeres how did the scaffolds located at termini behave during cell propagation? If the absence of telomere could explain some low read coverage scaffold, then hybrids that have been subject to many cell generation would have extremely low read coverage scaffolds which locate towards the end of synthetic microchromomes. We also

observed very few duck fragments inserted into the hamster genome, for which we speculated that it should be stable and have high read coverage.

Cause of the heterogeneity of hybrid cell line; we hypothesized that the scaffold read coverage could be classified into different read coverage state. Thus our first attempt was to use Hidden Markov Model (HMM) to model the read coverage using a sliding window of read counts along the scaffolds and thereby enabling to segment the scaffolds; we have tried 4 state-HMM in which state 1 meant absence, state 2 for shallow read coverage, state 3 for median read coverage and state 4 for high read coverage. For each state the read count was modeled by a binomial negative distribution. But the read coverage varied greatly within hybrids and sequencing depth varied among hybrids, which led to incorrect estimation of state in many cases. We have compared the false calling ratio between HMM and CBS, from which better segmentation was selected. Thus CBS had better performance and was therefore selected for our segmentation procedure and calling was determined by the mean value of each resultingsegment. With respect to the threshold set for CBS, great care was taken to avoid false calling. By considering our first estimation of expected read coverage for 20kb fragment and comparing the scaffolds which had been genotyped by Fluidigm qPCR technique, we found that most of the scaffolds with read coverage of 0.5 read per 20kb were considered as absent using the Fluidigm qPCR technique though few cases were present. For those low read coverage scaffolds, we checked the neighboring scaffolds using chicken coordinates and revealed that the neighboring scaffolds had similarity read coverage so that it would not cause false breakage. Furthermore, a buffering zone using read coverage between 0.5 and 1 read per 20kb fragment was set and called unknown (-). However, as mentioned previous in Result section, CBS required window size sometimes causes breakage imprecision or failure in detecting breakage. We therefore tried to devise our own segmentation procedure based on a pruned dynamic programming optimal change-point algorithm (G Riguail, ref to be completed) which does not require setting a window size. Unfortunately, this program cannot detect yet changes in read counts buried in the middle of largesegments and is therefore still under development..

Nevertheless, the “bugs” in CBS segmentation could be tolerated which was evident by two following reasons: (1) CBS segmentation allowed detecting most of the disrupted scaffolds and (2) the failure in detecting breakpoint resulted in shortening the distance of the markers on two ends of the same scaffolds but would not influence the following markers as they would be absent in most cases. Current RH maps made from CBS segmentation results

showed good agreements with the maps made by other genotyping techniques, reflecting that the segmentation should be robust though not perfect.

With new maps made for three chromosomes, we could therefore estimate the resolution and the power of this panel. The resolution had been defined above as the ratio of physical distance per cR and here we defined the power was that the mean minimum distance for observation of one breakage. For APL22, we calculated the resolution of different maps which were about 40kb/cR, 46kb/cR and 50kb/cR for APL22_GGA21, APL22_TGU21 and APL22_MGA23 respectively. The differences were owing to in each dataset contained different number of markers in which some were not shared by all references. However, comparative map of APL22_GGA21 by genotyping WGA-panel was 283cR in length consisting 24 markers (Rao et al. 2012), reflecting a resolution of 24.6kb/cR. This was a consequence of two different genotyping strategies as illustrated in Supplementary Figure7, the genotyping by sequencing only considered one breakage for scaffolds if there was at least one breakage which consequently led to shortening the actual distance on the map, whereas the conventional genotyping had opposite effect. Moreover, this significant difference in resolution for the same chromosome was evident in APL12. The length of APL12_FLDM made by genotyping non-WGA panel with Fluidigm qPCR was as two times long as that of APL12_GGA11, 727.5cR and 358.9cR respectively. Here we compared the resolution solely for chromosomes genotyped by sequencing. APL12_GGA11 showed a resolution of 58kb/cR while resolution of APL2_GGA2 was estimated to be 93.5kb/cR. Indeed, as illustrated above, the length of the map could be underestimated and therefore decreased the resolution. Hence we introduced the power (Θ) of the panel to estimate the frequency of the breakage which first demonstrated by Cox *et al* (Cox et al. 1990). Θ was estimated by the equation: $\Theta = (A^+B^- + A^-B^+)/[T(R_A + R_B - 2R_AR_B)]$ in which A^+B^- was observation that A was present B being absent (A^-B^+ was on the contrary), R was retention and T was the number of hybrids in the panel. When retention (r) reached 50%, the Θ could reach minimum. Θ estimated by this formula could be very independent on local retention; the purpose is to briefly estimate the interval to detect a breakage. Thus Θ was estimated to be 3.6cR for APL2 whereas Θ=2.5cR for APL12, which conversely reflected that observing one breakage on APL2 needed longer interval than that of APL12. Comparative mapping approach had been tested on pig RH data with about 5000 markers (B. Servin, unpublished data) and was successfully applied to validate the assembly of dog chromosome 2 (Servin et al. 2010).

In our approach, we adopted comparative mapping rather than traditional RH mapping had several advantages: (1) comparative mapping not only use multipoint likelihood but also integrate comparative data, so that the markers had very high LOD score that were very difficult to order by traditional RH mapping could be proposed by means of the reference genome; (2) for large number of hybrids, traditional RH mapping only can place a small proportion of marker on the framework map in which the order of marker was highly likely, whereas the comparative mapping usually gave more markers on robust map; and (3) comparative mapping was faster way to construct RH map especially when dealing with high throughput data. For instance, we tried traditional mapping on APL2 which contained more than 300 markers, only 56 of which were located on framework map using LOD score of 3 as a threshold, whereas 296 out of 308 markers were on robust map. This may infer that comparative mapping can somehow compensate the effect of bad genotyped markers.

Before our first attempt to using more than one species as reference, we were not confident that whether comparative mapping would give too much weight on reference genome and thereby the RH maps would be very different from different references. However the RH map of APL2 inferred that the RH maps were robust as they were highly consistent between APL2_GGA2 and APL2_TGU2 despite the extensive intrachromosomal rearrangements between chicken and zebra finch which diverged more than 100 million years (Pereira and Baker 2006). Furthermore, the different maps made with different references for the same chromosome they contained different set of markers, i.e. APL12, on APL12_GGA11 the order of 4 markers from sca2156_0 to sca5274 was not invariant in the map distribution during MCMC iterations, but was on robust map on APL12_TGU11. We suggested that we could integrate all those maps to increase the number of markers on the robust map.

All three duck chromosomes suggested that duck chromosomes experienced extensive intrachromosomal rearrangements since it diverged from its common ancestor with chicken, zebra finch and turkey. Unlike turkey whose major type of rearrangements was inversions compared to chicken, the rearrangements in duck were more complex as in zebra finch while comparing to chicken or turkey. Interestingly, the proposed centromeric regions were all involved in the rearrangements, but only with three chromosomes it would be difficult to hypothesize that (neo)centromeres play an important role in speciation. In addition, some regions showed that have the same order in chicken, zebra finch and turkey but were inverted in duck. On APL12, the region spanned from marker sca736_0 to sca903_0 was inverted

always in duck, meanwhile, two similar situations could be found on APL2 both of which involved 10Mb (seen in Figure 9, 10 and 11). It seemed that those inversions were prone to be duck-specific, but more evidences would be needed.

We also used the marker ordering of robust maps to investigate the evolutionary breakpoint regions for these three chromosomes. It is believed that evolutionary breakpoint share some common characteristics such as high GC-content, gene-rich or high repetitive content(Gordon et al. 2007). We took 5kb upstream and 5 kb downstream region surrounding the breakpoint while comparing duck RH map with other three birds, we assessed the GC content of those 5kb windows as well as the virtual chromosome made by concatenating all the scaffolds. Of the breakpoint regions on APL22, the GC content were relatively higher than genome average (about 41%, (Huang *et al, in prep*) (Supplementary Table5), however, the overall GC content for this microchromosome was high which was more than 45.1%. On the contrary, the breakpoint regions on APL2 showed an overall lower GC content than chromosome-wide even though that was 38.3%. Nevertheless, we did find that some scaffolds involved in breakpoint regions had repeat regardless of low GC content. Again, we only had limited data by far; to unveil more evidence to support the hypothesis more data from other chromosomes would be acquired. Additionally, those three chromosomes exhibited distinct isochore that the GC contents were 38.3%, 40.1% and 45.1% for APL2, APL12 and APL22 respectively.

Finally, Nacisse only aligned 1787 duck scaffolds onto chicken genome and we had usable data for 2027 scaffolds, meaning that there were more than 200 scaffolds were divergent from chicken or located on ChrUN like sca575 on APL12. Those scaffolds will be incorporated in the robust maps using their RH vectors.

Conclusion

We have sequenced the duck RH panel at a shallow sequencing depth, with a bulk of junk sequencing reads from hamster; we are still able to construct RH maps and thereby order and assign scaffolds onto duck chromosomes. We have compared the RH maps made by different genotyping methods, from which good consistence proves the feasibility of this survey study. Moreover, we used three references for comparative mapping duck scaffolds, from which the caveat that too much weight was posed on reference genome could be therefore eliminated and the maps will promote comparative studies for avian chromosome evolution. The maps we have represented above indicated extensive intrachromosomal

rearrangements which are not thoroughly understood from the available cytogenetic data (Fillon et al. 2007; Skinner et al. 2009). Although some procedures could be improved in the near future, our survey study provides an opportunity to overcome the shortage of NGS genome assemble by taking advantage of NGS technology.

Whole genome RH maps for duck are under construction; we believe that the availability of improved duck genome assembly will facilitate research in related field. Moreover, the comparative maps for all four sequenced birds will shed great light on avian chromosome evolution and reconstruction of ancestor genome.

Methods

Library preparation

The sequencing library was made according to manufacture's protocol (Illumina). Briefly, 1μg of genomic DNA was fragmented by sonication and size-selected by separation on agarose gel. Then the fragmented genomic DNA was polished and added an "A" base to the ends of the DNA fragments. DNA adaptors with a single "T" base overhang at the 3'end and a 6 nucleotides barcode for multiplexing were ligated to the above products. The mean insert size of the library was 335 bp.

Sequencing

One hundred hybrids were sequenced using an Illumina Hiseq2000 sequencing machine. For each hybrid 0.7 pg of DNA was used and twelve hybrids were multiplexed and sequenced in a single lane by pair-end sequencing, with a read length of 101 bases. Individual hybrids are identified by reading the barcode sequence on the adaptors.

Sequence Alignment and Data Filtering

As the hamster genome sequence is unavailable, the mouse genome was used as a reference to detect the donor cell sequence sequence reads. Alignment to the mouse genome was done with the GLINT alignment software (T.Faraut personnel communication). Alignment to the duck assembled scaffolds was done with the BWA alignment tools introduced by Li *et al* using default settings (Li and Durbin 2010). Only paired reads for which both sequences mapped at unique positions on duck scaffolds were retained for further analysis. However, reads that could be mapped both on the duck and the mouse genome were

discarded. After these filtering processes, new bam files were created containing only the paired reads uniquely mapped on duck scaffolds.

Scaffold Calling

To detect breakpoints along the scaffolds in the hybrids, the calling was done using the circular binary segmentation (CBS) algorithm introduced by Olshen et al. 2004 (Olshen et al. 2004), using a window size of 20kb. Fragments or ends of scaffolds not reaching 20kb were not included in the analysis. An output file was generated describing the segmentation of each hybrid, each segment being composed of windows with similar characteristics. Information describes the number of windows in a segment, its first and last window, the total number of reads it contains and the mean value for its 20kb windows. The mean number of sequencing reads for 20kb windows was used as the parameter to determine the genotype call: presence or absence of the scaffold segment in the hybrid. A Python script was used to summarize the scaffold calling for all 86 hybrids.

Map Construction

Draft maps (comprehensive maps) were made using the comparative mapping approach (Faraut et al. 2007) which is part of the Carthagene program (de Givry et al. 2005). Chicken, turkey and zebra finch genomes were used as references to build three sets of maps. First the RH vectors obtained by the scaffold calling and the files containing the ordering of the markers along the reference genomes were used to compute the marker ordering by 2-point likelihoods using the lkh command. Then the properties of the map posterior distributions were obtained with the mcmc command using 32806 as random generator seed and running 5000 mcmc iteration, the first 1000 of which were discarded. The output file from mcmc was used as input for the metamap program described by Servin *et al* (Servin et al. 2010), from which the robust map could be therefore obtained together with posterior possibility of each maps. Finally the RH map pictures for APL12 and APL22 were created using MapChart (Voorrips 2002). The view of the comparative maps of APL2 was made using an R script.

FISH experiments

Chicken BAC clones were chosen in the Wageningen BAC library according to their known position, as estimated by BAC end sequence information (Crooijmans et al. 2000), in regions paralogous to the breakpoint under study. WAG-21A17 (accession number

CZ567423.1) corresponds to marker sca713, WAG-15A21 (CZ561801) corresponds to sca616_0, while WAG-7I10 (CZ560582) correspond to sca1034_1 and WAG-23I13 (CZ568657) correspond to sca74_1. BAC clones were grown in LB medium with 12,5 µg/ml chloramphenicol. The DNA was extracted using the Qiagen plasmid midi kit.

FISH was carried out on metaphase spreads obtained from fibroblast cultures of 7-days old chicken and duck embryos, arrested with 0.05 µg/ml colcemid (Sigma) and fixed by standard procedures. The FISH protocol is derived from Yerle et al, 1992 (Yerle et al. 1992). Two-colour FISH was performed by labelling 100 ng for each BAC clones with alexa fluorochromes (ChromaTide® Alexa Fluor® 488-5-dUTP, Molecular probes; ChromaTide® Alexa Fluor® 568-5-dUTP, Molecular Probes) by random priming using the Bioprim Kit (Invitrogen). The probes were purified using spin column G50 Illustra (Amersham Biosciences). Probes were ethanol precipitated, resuspend in 50% formamide hybridization buffer (for FISH on chicken metaphases) or in 40% formamide hybridization buffer for heterologous FISH. Probes were hybridised to chicken metaphase slides for 17 hours at 37°C and to duck metaphases for 48H in the Hybridizer (Dako). Chromosomes were counterstained with DAPI in antifade solution (Vectashield with DAP, Vector). The hybridised metaphases were screened with a Zeiss fluorescence microscope and a minimum of twenty spreads was analysed for each experiment. Spot-bearing metaphases were captured and analysed with a cooled CCD camera using Cytovision software (Applied Imaging).

Reference

Abeel, T., T. Van Parys, Y. Saeys, J. Galagan, and Y. Van de Peer. 2012. GenomeView: a next-generation genome browser. *Nucleic Acids Res* 40: e12.

Benham, F., K. Hart, J. Crolla, M. Bobrow, M. Francavilla, and P.N. Goodfellow. 1989. A method for generating hybrids containing nonselected fragments of human chromosomes. *Genomics* 4: 509-517.

Breen, M., C. Hitte, T.D. Lorentzen, R. Thomas, E. Cadieu, L. Sabacan, A. Scott, G. Evanno, H.G. Parker, E.F. Kirkness, R. Hudson, R. Guyon, G.G. Mahairas, B. Gelfenbeyn, C.M. Fraser, C. Andre, F. Galibert, and E.A. Ostrander. 2004. An integrated 4249 marker FISH/RH map of the canine genome. *BMC Genomics* 5: 65.

Breen, M., S. Jouquand, C. Renier, C.S. Mellersh, C. Hitte, N.G. Holmes, A. Cheron, N. Suter, F. Vignaux, A.E. Bristow, C. Priat, E. McCann, C. Andre, S. Boundy, P. Gitsham, R.

Thomas, W.L. Bridge, H.F. Spriggs, E.J. Ryder, A. Curson, J. Sampson, E.A. Ostrander, M.M. Binns, and F. Galibert. 2001. Chromosome-specific single-locus FISH probes allow anchorage of an 1800-marker integrated radiation-hybrid/linkage map of the domestic dog genome to all chromosomes. *Genome Res* 11: 1784-1795.

Carroll, S.M., M.L. DeRose, P. Gaudray, C.M. Moore, D.R. Needham-Vandevanter, D.D. Von Hoff, and G.M. Wahl. 1988. Double minute chromosomes can be produced from precursors derived from a chromosomal deletion. *Mol Cell Biol* 8: 1525-1533.

Carroll, S.M., P. Gaudray, M.L. De Rose, J.F. Emery, J.L. Meinkoth, E. Nakkim, M. Subler, D.D. Von Hoff, and G.M. Wahl. 1987. Characterization of an episome produced in hamster cells that amplify a transfected CAD gene at high frequency: functional evidence for a mammalian replication origin. *Mol Cell Biol* 7: 1740-1750.

Courcelle, E., Y. Beausse, S. Letort, O. Stahl, R. Fremez, C. Ngom-Bru, J. Gouzy, and T. Faraut. 2008. Narcisse: a mirror view of conserved syntenies. *Nucleic Acids Res* 36: D485-490.

Cox, D.R., M. Burmeister, E.R. Price, S. Kim, and R.M. Myers. 1990. Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. *Science* 250: 245-250.

Crooijmans, R.P., J. Vrebalov, R.J. Dijkhof, J.J. van der Poel, and M.A. Groenen. 2000. Two-dimensional screening of the Wageningen chicken BAC library. *Mamm Genome* 11: 360-363.

Dalloul, R.A., J.A. Long, A.V. Zimin, L. Aslam, K. Beal, A. Blomberg Le, P. Bouffard, D.W. Burt, O. Crasta, R.P. Crooijmans, K. Cooper, R.A. Coulombe, S. De, M.E. Delany, J.B. Dodgson, J.J. Dong, C. Evans, K.M. Frederickson, P. Flicek, L. Florea, O. Folkerts, M.A. Groenen, T.T. Harkins, J. Herrero, S. Hoffmann, H.J. Megens, A. Jiang, P. de Jong, P. Kaiser, H. Kim, K.W. Kim, S. Kim, D. Langenberger, M.K. Lee, T. Lee, S. Mane, G. Marcais, M. Marz, A.P. McElroy, T. Modise, M. Nefedov, C. Notredame, I.R. Paton, W.S. Payne, G. Pertea, D. Prickett, D. Puiu, D. Qioa, E. Raineri, M. Ruffier, S.L. Salzberg, M.C. Schatz, C. Scheuring, C.J. Schmidt, S. Schroeder, S.M. Searle, E.J. Smith, J. Smith, T.S. Sonstegard, P.F. Stadler, H. Tafer, Z.J. Tu, C.P. Van Tassell, A.J. Vilella, K.P. Williams, J.A. Yorke, L. Zhang, H.B. Zhang, X. Zhang, Y. Zhang, and K.M. Reed. 2010. Multi-platform next-generation sequencing of the domestic turkey (Meleagris gallopavo): genome assembly and analysis. *PLoS Biol* 8.

de Givry, S., M. Bouchez, P. Chabrier, D. Milan, and T. Schiex. 2005. CARHTA GENE: multipopulation integrated genetic and radiation hybrid mapping. *Bioinformatics* 21: 1703-1704.

Elsik, C.G. R.L. Tellam K.C. Worley R.A. Gibbs D.M. Muzny G.M. Weinstock D.L. Adelson E.E. Eichler L. Elnitski R. Guigo D.L. Hamernik S.M. Kappes H.A. Lewin D.J. Lynn F.W. Nicholas A. Reymond M. Rijnkels L.C. Skow E.M. Zdobnov L. Schook J. Womack T. Alioto S.E. Antonarakis A. Astashyn C.E. Chapple H.C. Chen J. Chrast F. Camara O. Ermolaeva C.N. Henrichsen W. Hlavina Y. Kapustin B. Kiryutin P. Kitts F. Kokocinski M. Landrum D. Maglott K. Pruitt V. Sapojnikov S.M. Searle V. Solovyev A. Souvorov C. Ucla C. Wyss J.M. Anzola D. Gerlach E. Elhaik D. Graur J.T. Reese R.C. Edgar J.C. McEwan G.M. Payne J.M. Raison T. Junier E.V. Kriventseva E. Eyras M. Plass R. Donthu D.M. Larkin J. Reecy M.Q. Yang L. Chen Z. Cheng C.G. Chitko-McKown G.E. Liu L.K. Matukumalli J. Song B. Zhu D.G. Bradley F.S. Brinkman L.P. Lau M.D. Whiteside A. Walker T.T. Wheeler T. Casey J.B. German D.G. Lemay N.J. Maqbool A.J. Molenaar S. Seo P. Stothard C.L. Baldwin R. Baxter C.L. Brinkmeyer-Langford W.C. Brown C.P. Childers T. Connelley S.A. Ellis K. Fritz E.J. Glass C.T. Herzig A. Iivanainen K.K. Lahmers A.K. Bennett C.M. Dickens J.G. Gilbert D.E. Hagen H. Salih J. Aerts A.R. Caetano B. Dalrymple J.F. Garcia C.A. Gill S.G. Hiendleder E. Memili D. Spurlock J.L. Williams L. Alexander M.J. Brownstein L. Guan R.A. Holt S.J. Jones M.A. Marra R. Moore S.S. Moore A. Roberts M. Taniguchi R.C. Waterman J. Chacko M.M. Chandrabose A. Cree M.D. Dao H.H. Dinh R.A. Gabisi S. Hines J. Hume S.N. Jhangiani V. Joshi C.L. Kovar L.R. Lewis Y.S. Liu J. Lopez M.B. Morgan N.B. Nguyen G.O. Okwuonu S.J. Ruiz J. Santibanez R.A. Wright C. Buhay Y. Ding S. Dugan-Rocha J. Herdandez M. Holder A. Sabo A. Egan J. Goodell K. Wilczek-Boney G.R. Fowler M.E. Hitchens R.J. Lozado C. Moen D. Steffen J.T. Warren J. Zhang R. Chiu J.E. Schein K.J. Durbin P. Havlak H. Jiang Y. Liu X. Qin Y. Ren Y. Shen H. Song S.N. Bell C. Davis A.J. Johnson S. Lee L.V. Nazareth B.M. Patel L.L. Pu S. Vattathil R.L. Williams, Jr. S. Curry C. Hamilton E. Sodergren D.A. Wheeler W. Barris G.L. Bennett A. Eggen R.D. Green G.P. Harhay M. Hobbs O. Jann J.W. Keele M.P. Kent S. Lien S.D. McKay S. McWilliam A. Ratnakumar R.D. Schnabel T. Smith W.M. Snelling T.S. Sonstegard R.T. Stone Y. Sugimoto A. Takasuga J.F. Taylor C.P. Van Tassell M.D. Macneil A.R. Abatepaulo C.A. Abbey V. Ahola I.G. Almeida A.F. Amadio E. Anatriello S.M. Bahadue F.H. Biase C.R. Boldt J.A. Carroll W.A. Carvalho E.P. Cervelatti E. Chacko J.E. Chapin Y.

Cheng J. Choi A.J. Colley T.A. de Campos M. De Donato I.K. Santos C.J. de Oliveira H. Deobald E. Devinoy K.E. Donohue P. Dovc A. Eberlein C.J. Fitzsimmons A.M. Franzin G.R. Garcia S. Genini C.J. Gladney J.R. Grant M.L. Greaser J.A. Green D.L. Hadsell H.A. Hakimov R. Halgren J.L. Harrow E.A. Hart N. Hastings M. Hernandez Z.L. Hu A. Ingham T. Iso-Touru C. Jamis K. Jensen D. Kapetis T. Kerr S.S. Khalil H. Khatib D. Kolbehdari C.G. Kumar D. Kumar R. Leach J.C. Lee C. Li K.M. Logan R. Malinverni E. Marques W.F. Martin N.F. Martins S.R. Maruyama R. Mazza K.L. McLean J.F. Medrano B.T. Moreno D.D. More C.T. Muntean H.P. Nandakumar M.F. Nogueira I. Olsaker S.D. Pant F. Panzitta R.C. Pastor M.A. Poli N. Poslusny S. Rachagani S. Ranganathan A. Razpet P.K. Riggs G. Rincon N. Rodriguez-Osorio S.L. Rodriguez-Zas N.E. Romero A. Rosenwald L. Sando S.M. Schmutz L. Shen L. Sherman B.R. Southey Y.S. Lutzow J.V. Sweedler I. Tammen B.P. Telugu J.M. Urbanski Y.T. Utsunomiya C.P. Verschoor A.J. Waardenberg Z. Wang R. Ward R. Weikard T.H. Welsh, Jr. S.N. White L.G. Wilming K.R. Wunderlich J. Yang and F.Q. Zhao. 2009. The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science* 324: 522-528.

Faraut, T., S. de Givry, P. Chabrier, T. Derrien, F. Galibert, C. Hitte, and T. Schiex. 2007. A comparative genome approach to marker ordering. *Bioinformatics* 23: e50-56.

Fillon, V., M. Vignoles, R.P. Crooijmans, M.A. Groenen, R. Zoorob, and A. Vignal. 2007. FISH mapping of 57 BAC clones reveals strong conservation of synteny between Galliformes and Anseriformes. *Anim Genet* 38: 303-307.

Gibbs, R.A. G.M. Weinstock M.L. Metzker D.M. Muzny E.J. Sodergren S. Scherer G. Scott D. Steffen K.C. Worley P.E. Burch G. Okwuonu S. Hines L. Lewis C. DeRamo O. Delgado S. Dugan-Rocha G. Miner M. Morgan A. Hawes R. Gill Celera R.A. Holt M.D. Adams P.G. Amanatides H. Baden-Tillson M. Barnstead S. Chin C.A. Evans S. Ferriera C. Fosler A. Glodek Z. Gu D. Jennings C.L. Kraft T. Nguyen C.M. Pfannkoch C. Sitter G.G. Sutton J.C. Venter T. Woodage D. Smith H.M. Lee E. Gustafson P. Cahill A. Kana L. Doucette-Stamm K. Weinstock K. Fechtel R.B. Weiss D.M. Dunn E.D. Green R.W. Blakesley G.G. Bouffard P.J. De Jong K. Osoegawa B. Zhu M. Marra J. Schein I. Bosdet C. Fjell S. Jones M. Krzywinski C. Mathewson A. Siddiqui N. Wye J. McPherson S. Zhao C.M. Fraser J. Shetty S. Shatsman K. Geer Y. Chen S. Abramzon W.C. Nierman P.H. Havlak R. Chen K.J. Durbin A. Egan Y. Ren X.Z. Song B. Li Y. Liu X. Qin S. Cawley A.J. Cooney L.M. D'Souza K. Martin J.Q. Wu M.L. Gonzalez-Garay A.R. Jackson K.J. Kalafus M.P. McLeod A. Milosavljevic

668 D. Virk A. Volkov D.A. Wheeler Z. Zhang J.A. Bailey E.E. Eichler E. Tuzun E.
669 Birney E. Mongin A. Ureta-Vidal C. Woodwark E. Zdobnov P. Bork M. Suyama D.
670 Torrents M. Alexandersson B.J. Trask J.M. Young H. Huang H. Wang H. Xing S.
671 Daniels D. Gietzen J. Schmidt K. Stevens U. Vitt J. Wingrove F. Camara M. Mar Alba
672 J.F. Abril R. Guigo A. Smit I. Dubchak E.M. Rubin O. Couronne A. Poliakov N.
673 Hubner D. Ganten C. Goesele O. Hummel T. Kreitler Y.A. Lee J. Monti H. Schulz H.
674 Zimdahl H. Himmelbauer H. Lehrach H.J. Jacob S. Bromberg J. Gullings-Handley
675 M.I. Jensen-Seaman A.E. Kwitek J. Lazar D. Pasko P.J. Tonellato S. Twigger C.P.
676 Ponting J.M. Duarte S. Rice L. Goodstadt S.A. Beatson R.D. Emes E.E. Winter C.
677 Webber P. Brandt G. Nyakatura M. Adetobi F. Chiaromonte L. Elnitski P. Eswara R.C.
678 Hardison M. Hou D. Kolbe K. Makova W. Miller A. Nekrutenko C. Riemer S.
679 Schwartz J. Taylor S. Yang Y. Zhang K. Lindpaintner T.D. Andrews M. Caccamo M.
680 Clamp L. Clarke V. Curwen R. Durbin E. Eyras S.M. Searle G.M. Cooper S.
681 Batzoglou M. Brudno A. Sidow E.A. Stone B.A. Payseur G. Bourque C. Lopez-Otin
682 X.S. Puente K. Chakrabarti S. Chatterji C. Dewey L. Pachter N. Bray V.B. Yap A.
683 Caspi G. Tesler P.A. Pevzner D. Haussler K.M. Roskin R. Baertsch H. Clawson T.S.
684 Furey A.S. Hinrichs D. Karolchik W.J. Kent K.R. Rosenbloom H. Trumbower M.
685 Weirauch D.N. Cooper P.D. Stenson B. Ma M. Brent M. Arumugam D. Shteynberg
686 R.R. Copley M.S. Taylor H. Riethman U. Mudunuri J. Peterson M. Guyer A.
687 Felsenfeld S. Old S. Mockrin and F. Collins. 2004. Genome sequence of the Brown
688 Norway rat yields insights into mammalian evolution. *Nature* 428: 493-521.

689 Goodfellow, P.J., S. Povey, H.A. Nevanlinna, and P.N. Goodfellow. 1990. Generation of a
690 panel of somatic cell hybrids containing unselected fragments of human chromosome
691 10 by X-ray irradiation and cell fusion: application to isolating the MEN2A region in
692 hybrid cells. *Somat Cell Mol Genet* 16: 163-171.

693 Gordon, L., S. Yang, M. Tran-Gyamfi, D. Baggott, M. Christensen, A. Hamilton, R.
694 Crooijmans, M. Groenen, S. Lucas, I. Ovcharenko, and L. Stubbs. 2007. Comparative
695 analysis of chicken chromosome 28 provides new clues to the evolutionary fragility of
696 gene-rich vertebrate regions. *Genome Res* 17: 1603-1613.

697 Green, E.D. 2001. Strategies for the systematic sequencing of complex genomes. *Nat Rev*
698 *Genet* 2: 573-583.

699 Huang, Y. and I.d.g.s. consortium. 2012. The duck genome and transcriptome provide insight
700 into a viral-reservoir species.

701 Huang, Y., Y. Zhao, C.S. Haley, S. Hu, J. Hao, C. Wu, and N. Li. 2006. A genetic and
702 cytogenetic map for the duck (Anas platyrhynchos). *Genetics* 173: 287-296.
703 ICGC. 2004. Sequence and comparative analysis of the chicken genome provide unique
704 perspectives on vertebrate evolution. *Nature* 432: 695-716.
705 Lander, E.S. L.M. Linton B. Birren C. Nusbaum M.C. Zody J. Baldwin K. Devon K. Dewar
706 M. Doyle W. FitzHugh R. Funke D. Gage K. Harris A. Heaford J. Howland L. Kann J.
707 Lehoczky R. LeVine P. McEwan K. McKernan J. Meldrim J.P. Mesirov C. Miranda
708 W. Morris J. Naylor C. Raymond M. Rosetti R. Santos A. Sheridan C. Sougnez N.
709 Stange-Thomann N. Stojanovic A. Subramanian D. Wyman J. Rogers J. Sulston R.
710 Ainscough S. Beck D. Bentley J. Burton C. Clee N. Carter A. Coulson R. Deadman P.
711 Deloukas A. Dunham I. Dunham R. Durbin L. French D. Grafham S. Gregory T.
712 Hubbard S. Humphray A. Hunt M. Jones C. Lloyd A. McMurray L. Matthews S.
713 Mercer S. Milne J.C. Mullikin A. Mungall R. Plumb M. Ross R. Shownkeen S. Sims
714 R.H. Waterston R.K. Wilson L.W. Hillier J.D. McPherson M.A. Marra E.R. Mardis
715 L.A. Fulton A.T. Chinwalla K.H. Pepin W.R. Gish S.L. Chissoe M.C. Wendl K.D.
716 Delehaunty T.L. Miner A. Delehaunty J.B. Kramer L.L. Cook R.S. Fulton D.L.
717 Johnson P.J. Minx S.W. Clifton T. Hawkins E. Branscomb P. Predki P. Richardson S.
718 Wenning T. Slezak N. Doggett J.F. Cheng A. Olsen S. Lucas C. Elkin E. Uberbacher
719 M. Frazier R.A. Gibbs D.M. Muzny S.E. Scherer J.B. Bouck E.J. Sodergren K.C.
720 Worley C.M. Rives J.H. Gorrell M.L. Metzker S.L. Naylor R.S. Kucherlapati D.L.
721 Nelson G.M. Weinstock Y. Sakaki A. Fujiyama M. Hattori T. Yada A. Toyoda T. Itoh
722 C. Kawagoe H. Watanabe Y. Totoki T. Taylor J. Weissenbach R. Heilig W. Saurin F.
723 Artiguenave P. Brottier T. Bruls E. Pelletier C. Robert P. Wincker D.R. Smith L.
724 Doucette-Stamm M. Rubenfield K. Weinstock H.M. Lee J. Dubois A. Rosenthal M.
725 Platzer G. Nyakatura S. Taudien A. Rump H. Yang J. Yu J. Wang G. Huang J. Gu L.
726 Hood L. Rowen A. Madan S. Qin R.W. Davis N.A. Federspiel A.P. Abola M.J.
727 Proctor R.M. Myers J. Schmutz M. Dickson J. Grimwood D.R. Cox M.V. Olson R.
728 Kaul N. Shimizu K. Kawasaki S. Minoshima G.A. Evans M. Athanasiou R. Schultz
729 B.A. Roe F. Chen H. Pan J. Ramser H. Lehrach R. Reinhardt W.R. McCombie M. de
730 la Bastide N. Dedhia H. Blocker K. Hornischer G. Nordsiek R. Agarwala L. Aravind
731 J.A. Bailey A. Bateman S. Batzoglou E. Birney P. Bork D.G. Brown C.B. Burge L.
732 Cerutti H.C. Chen D. Church M. Clamp R.R. Copley T. Doerks S.R. Eddy E.E.
733 Eichler T.S. Furey J. Galagan J.G. Gilbert C. Harmon Y. Hayashizaki D. Haussler H.
734 Hermjakob K. Hokamp W. Jang L.S. Johnson T.A. Jones S. Kasif A. Kaspryzk S.

Kennedy W.J. Kent P. Kitts E.V. Koonin I. Korf D. Kulp D. Lancet T.M. Lowe A. McLysaght T. Mikkelsen J.V. Moran N. Mulder V.J. Pollara C.P. Ponting G. Schuler J. Schultz G. Slater A.F. Smit E. Stupka J. Szustakowski D. Thierry-Mieg J. Thierry-Mieg L. Wagner J. Wallis R. Wheeler A. Williams Y.I. Wolf K.H. Wolfe S.P. Yang R.F. Yeh F. Collins M.S. Guyer J. Peterson A. Felsenfeld K.A. Wetterstrand A. Patrinos M.J. Morgan P. de Jong J.J. Catanese K. Osoegawa H. Shizuya S. Choi and Y.J. Chen. 2001. Initial sequencing and analysis of the human genome. *Nature* 409: 860-921.

Lewin, H.A., D.M. Larkin, J. Pontius, and S.J. O'Brien. 2009. Every genome sequence needs a good map. *Genome Res* 19: 1925-1928.

Li, H. and R. Durbin. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26: 589-595.

Li, R. W. Fan G. Tian H. Zhu L. He J. Cai Q. Huang Q. Cai B. Li Y. Bai Z. Zhang Y. Zhang W. Wang J. Li F. Wei H. Li M. Jian R. Nielsen D. Li W. Gu Z. Yang Z. Xuan O.A. Ryder F.C. Leung Y. Zhou J. Cao X. Sun Y. Fu X. Fang X. Guo B. Wang R. Hou F. Shen B. Mu P. Ni R. Lin W. Qian G. Wang C. Yu W. Nie J. Wang Z. Wu H. Liang J. Min Q. Wu S. Cheng J. Ruan M. Wang Z. Shi M. Wen B. Liu X. Ren H. Zheng D. Dong K. Cook G. Shan H. Zhang C. Kosiol X. Xie Z. Lu Y. Li C.C. Steiner T.T. Lam S. Lin Q. Zhang G. Li J. Tian T. Gong H. Liu D. Zhang L. Fang C. Ye J. Zhang W. Hu A. Xu Y. Ren G. Zhang M.W. Bruford Q. Li L. Ma Y. Guo N. An Y. Hu Y. Zheng Y. Shi Z. Li Q. Liu Y. Chen J. Zhao N. Qu S. Zhao F. Tian X. Wang H. Wang L. Xu X. Liu T. Vinar Y. Wang T.W. Lam S.M. Yiu S. Liu Y. Huang G. Yang Z. Jiang N. Qin L. Li L. Bolund K. Kristiansen G.K. Wong M. Olson X. Zhang S. Li and H. Yang. 2010. The sequence and de novo assembly of the giant panda genome. *Nature* 463: 311-317.

Lindblad-Toh, K. C.M. Wade T.S. Mikkelsen E.K. Karlsson D.B. Jaffe M. Kamal M. Clamp J.L. Chang E.J. Kulbokas, 3rd M.C. Zody E. Mauceli X. Xie M. Breen R.K. Wayne E.A. Ostrander C.P. Ponting F. Galibert D.R. Smith P.J. DeJong E. Kirkness P. Alvarez T. Biagi W. Brockman J. Butler C.W. Chin A. Cook J. Cuff M.J. Daly D. DeCaprio S. Gnerre M. Grabherr M. Kellis M. Kleber C. Bardeleben L. Goodstadt A. Heger C. Hitte L. Kim K.P. Koepfli H.G. Parker J.P. Pollinger S.M. Searle N.B. Sutter R. Thomas C. Webber J. Baldwin A. Abebe A. Abouelleil L. Aftuck M. Ait-Zahra T. Aldredge N. Allen P. An S. Anderson C. Antoine H. Arachchi A. Aslam L. Ayotte P. Bachantsang A. Barry T. Bayul M. Benamara A. Berlin D. Bessette B. Blitshteyn T.

769 Bloom J. Blye L. Boguslavskiy C. Bonnet B. Boukhgalter A. Brown P. Cahill N.
770 Calixte J. Camarata Y. Cheshatsang J. Chu M. Citroen A. Collymore P. Cooke T.
771 Dawoe R. Daza K. Decktor S. DeGray N. Dhargay K. Dooley P. Dorje K. Dorjee L.
772 Dorris N. Duffey A. Dupes O. Egbiremolen R. Elong J. Falk A. Farina S. Faro D.
773 Ferguson P. Ferreira S. Fisher M. FitzGerald K. Foley C. Foley A. Franke D. Friedrich
774 D. Gage M. Garber G. Gearin G. Giannoukos T. Goode A. Goyette J. Graham E.
775 Grandbois K. Gyaltsen N. Hafez D. Hagopian B. Hagos J. Hall C. Healy R. Hegarty T.
776 Honan A. Horn N. Houde L. Hughes L. Hunnicutt M. Husby B. Jester C. Jones A.
777 Kamat B. Kanga C. Kells D. Khazanovich A.C. Kieu P. Kisner M. Kumar K. Lance T.
778 Landers M. Lara W. Lee J.P. Leger N. Lennon L. Leuper S. LeVine J. Liu X. Liu Y.
779 Lokyitsang T. Lokyitsang A. Lui J. Macdonald J. Major R. Marabella K. Maru C.
780 Matthews S. McDonough T. Mehta J. Meldrim A. Melnikov L. Meneus A. Mihalev T.
781 Mihova K. Miller R. Mittelman V. Mlenga L. Mulrain G. Munson A. Navidi J. Naylor
782 T. Nguyen N. Nguyen C. Nguyen R. Nicol N. Norbu C. Norbu N. Novod T. Nyima P.
783 Olandt B. O'Neill K. O'Neill S. Osman L. Oyono C. Patti D. Perrin P. Phunkhang F.
784 Pierre M. Priest A. Rachupka S. Raghuraman R. Rameau V. Ray C. Raymond F. Rege
785 C. Rise J. Rogers P. Rogov J. Sahalie S. Settipalli T. Sharpe T. Shea M. Sheehan N.
786 Sherpa J. Shi D. Shih J. Sloan C. Smith T. Sparrow J. Stalker N. Stange-Thomann S.
787 Stavropoulos C. Stone S. Stone S. Sykes P. Tchuinga P. Tenzing S. Tesfaye D.
788 Thoulutsang Y. Thoulutsang K. Topham I. Topping T. Tsamla H. Vassiliev V.
789 Venkataraman A. Vo T. Wangchuk T. Wangdi M. Weiand J. Wilkinson A. Wilson S.
790 Yadav S. Yang X. Yang G. Young Q. Yu J. Zainoun L. Zembek A. Zimmer and E.S.
791 Lander. 2005. Genome sequence, comparative analysis and haplotype structure of the
792 domestic dog. *Nature* 438: 803-819.

793 Morisson, M., A. Lemiere, S. Bosc, M. Galan, F. Plisson-Petit, P. Pinton, C. Delcros, K. Feve,
794 F. Pitel, V. Fillon, M. Yerle, and A. Vignal. 2002. ChickRH6: a chicken whole-
795 genome radiation hybrid panel. *Genet Sel Evol* 34: 521-533.

796 Olshen, A.B., E.S. Venkatraman, R. Lucito, and M. Wigler. 2004. Circular binary
797 segmentation for the analysis of array-based DNA copy number data. *Biostatistics* 5:
798 557-572.

799 Pereira, S.L. and A.J. Baker. 2006. A mitogenomic timescale for birds detects variable
800 phylogenetic rates of molecular evolution and refutes the standard molecular clock.
801 *Mol Biol Evol* 23: 1731-1740.

802 Pontius, J.U., J.C. Mullikin, D.R. Smith, K. Lindblad-Toh, S. Gnerre, M. Clamp, J. Chang, R.
803 Stephens, B. Neelam, N. Volfovsky, A.A. Schaffer, R. Agarwala, K. Narfstrom, W.J.
804 Murphy, U. Giger, A.L. Roca, A. Antunes, M. Menotti-Raymond, N. Yuhki, J. Pecon-
805 Slattery, W.E. Johnson, G. Bourque, G. Tesler, and S.J. O'Brien. 2007. Initial
806 sequence and comparative analysis of the cat genome. *Genome Res* 17: 1675-1689.
807 Qiu, Q., G. Zhang, T. Ma, W. Qian, J. Wang, Z. Ye, C. Cao, Q. Hu, J. Kim, D.M. Larkin, L.
808 Auvil, B. Capitanu, J. Ma, H.A. Lewin, X. Qian, Y. Lang, R. Zhou, L. Wang, K.
809 Wang, J. Xia, S. Liao, S. Pan, X. Lu, H. Hou, Y. Wang, X. Zang, Y. Yin, H. Ma, J.
810 Zhang, Z. Wang, Y. Zhang, D. Zhang, T. Yonezawa, M. Hasegawa, Y. Zhong, W. Liu,
811 Z. Huang, S. Zhang, R. Long, H. Yang, J.A. Lenstra, D.N. Cooper, Y. Wu, P. Shi, and
812 J. Liu. 2012. The yak genome and adaptation to life at high altitude. *Nat Genet* 44:
813 946-949.
814 Rao, M., M. Morisson, T. Faraut, S. Bardes, K. Feve, E. Labarthe, Y. Huang, N. Li, and A.
815 Vignal. 2012. A duck RH panel and its potential for assisting NGS genome assembly.
816 Schimke, R.T. 1984. Gene amplification in cultured animal cells. *Cell* 37: 705-713.
817 Sen, A. and M.S. Srivastava. 1975. On tests for detecting change in mean. *Ann. Statist* 3: 98-
818 108.
819 Servin, B., S. de Givry, and T. Faraut. 2010. Statistical confidence measures for genome maps:
820 application to the validation of genome assemblies. *Bioinformatics* 26: 3035-3042.
821 Skinner, B.M., L.B. Robertson, H.G. Tempest, E.J. Langley, D. Ioannou, K.E. Fowler, R.P.
822 Crooijmans, A.D. Hall, D.K. Griffin, and M. Volker. 2009. Comparative genomics in
823 chicken and Pekin duck using FISH mapping and microarray analysis. *BMC Genomics*
824 10: 357.
825 Stark, G.R. 1986. DNA amplification in drug resistant cells and in tumours. *Cancer Surv* 5: 1-
826 23.
827 Voorrips, R.E. 2002. MapChart: software for the graphical presentation of linkage maps and
828 QTLs. *J Hered* 93: 77-78.
829 Warren, W.C., D.F. Clayton, H. Ellegren, A.P. Arnold, L.W. Hillier, A. Kunstner, S. Searle, S.
830 White, A.J. Vilella, S. Fairley, A. Heger, L. Kong, C.P. Ponting, E.D. Jarvis, C.V.
831 Mello, P. Minx, P. Lovell, T.A. Velho, M. Ferris, C.N. Balakrishnan, S. Sinha, C.
832 Blatti, S.E. London, Y. Li, Y.C. Lin, J. George, J. Sweedler, B. Southey, P. Gunaratne,
833 M. Watson, K. Nam, N. Backstrom, L. Smeds, B. Nabholz, Y. Itoh, O. Whitney, A.R.
834 Pfenning, J. Howard, M. Volker, B.M. Skinner, D.K. Griffin, L. Ye, W.M. McLaren,
835 P. Flicek, V. Quesada, G. Velasco, C. Lopez-Otin, X.S. Puente, T. Olender, D. Lancet,

A.F. Smit, R. Hubley, M.K. Konkel, J.A. Walker, M.A. Batzer, W. Gu, D.D. Pollock, L. Chen, Z. Cheng, E.E. Eichler, J. Stapley, J. Slate, R. Ekblom, T. Birkhead, T. Burke, D. Burt, C. Scharff, I. Adam, H. Richard, M. Sultan, A. Soldatov, H. Lehrach, S.V. Edwards, S.P. Yang, X. Li, T. Graves, L. Fulton, J. Nelson, A. Chinwalla, S. Hou, E.R. Mardis, and R.K. Wilson. 2010. The genome of a songbird. *Nature* 464: 757-762.

Waterston, R.H. K. Lindblad-Toh E. Birney J. Rogers J.F. Abril P. Agarwal R. Agarwala R. Ainscough M. Alexandersson P. An S.E. Antonarakis J. Attwood R. Baertsch J. Bailey K. Barlow S. Beck E. Berry B. Birren T. Bloom P. Bork M. Botcherby N. Bray M.R. Brent D.G. Brown S.D. Brown C. Bult J. Burton J. Butler R.D. Campbell P. Carninci S. Cawley F. Chiaromonte A.T. Chinwalla D.M. Church M. Clamp C. Clee F.S. Collins L.L. Cook R.R. Copley A. Coulson O. Couronne J. Cuff V. Curwen T. Cutts M. Daly R. David J. Davies K.D. Delehaunty J. Deri E.T. Dermitzakis C. Dewey N.J. Dickens M. Diekhans S. Dodge I. Dubchak D.M. Dunn S.R. Eddy L. Elnitski R.D. Emes P. Eswara E. Eyras A. Felsenfeld G.A. Fewell P. Flicek K. Foley W.N. Frankel L.A. Fulton R.S. Fulton T.S. Furey D. Gage R.A. Gibbs G. Glusman S. Gnerre N. Goldman L. Goodstadt D. Grafham T.A. Graves E.D. Green S. Gregory R. Guigo M. Guyer R.C. Hardison D. Haussler Y. Hayashizaki L.W. Hillier A. Hinrichs W. Hlavina T. Holzer F. Hsu A. Hua T. Hubbard A. Hunt I. Jackson D.B. Jaffe L.S. Johnson M. Jones T.A. Jones A. Joy M. Kamal E.K. Karlsson D. Karolchik A. Kasprzyk J. Kawai E. Keibler C. Kells W.J. Kent A. Kirby D.L. Kolbe I. Korf R.S. Kucherlapati E.J. Kulbokas D. Kulp T. Landers J.P. Leger S. Leonard I. Letunic R. Levine J. Li M. Li C. Lloyd S. Lucas B. Ma D.R. Maglott E.R. Mardis L. Matthews E. Mauceli J.H. Mayer M. McCarthy W.R. McCombie S. McLaren K. McLay J.D. McPherson J. Meldrim B. Meredith J.P. Mesirov W. Miller T.L. Miner E. Mongin K.T. Montgomery M. Morgan R. Mott J.C. Mullikin D.M. Muzny W.E. Nash J.O. Nelson M.N. Nhan R. Nicol Z. Ning C. Nusbaum M.J. O'Connor Y. Okazaki K. Oliver E. Overton-Larty L. Pachter G. Parra K.H. Pepin J. Peterson P. Pevzner R. Plumb C.S. Pohl A. Poliakov T.C. Ponce C.P. Ponting S. Potter M. Quail A. Reymond B.A. Roe K.M. Roskin E.M. Rubin A.G. Rust R. Santos V. Sapojnikov B. Schultz J. Schultz M.S. Schwartz S. Schwartz C. Scott S. Seaman S. Searle T. Sharpe A. Sheridan R. Shownkeen S. Sims J.B. Singer G. Slater A. Smit D.R. Smith B. Spencer A. Stabenau N. Stange-Thomann C. Sugnet M. Suyama G. Tesler J. Thompson D. Torrents E. Trevaskis J. Tromp C. Ucla A. Ureta-Vidal J.P. Vinson A.C. Von Niederhausern C.M. Wade M. Wall R.J. Weber R.B. Weiss M.C. Wendl A.P. West K. Wetterstrand R.

Wheeler S. Whelan J. Wierzbowski D. Willey S. Williams R.K. Wilson E. Winter K.C. Worley D. Wyman S. Yang S.P. Yang E.M. Zdobnov M.C. Zody and E.S. Lander. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* 420: 520-562.

Yerle, M., O. Galman, Y. Lahbib-Mansais, and J. Gellin. 1992. Localization of the pig luteinizing hormone/choriogonadotropin receptor gene (LHCGR) by radioactive and nonradioactive in situ hybridization. *Cytogenetics and Cell Genetics* 59: 48-51.

Zhang, Y., X. Zhang, T.H. O'Hare, W.S. Payne, J.J. Dong, C.F. Scheuring, M. Zhang, J.J. Huang, M.K. Lee, M.E. Delany, H.B. Zhang, and J.B. Dodgson. 2012. A comparative physical map reveals the pattern of chromosomal evolution between the turkey (Meleagris gallopavo) and chicken (Gallus gallus) genomes. *BMC Genomics* 12: 447.

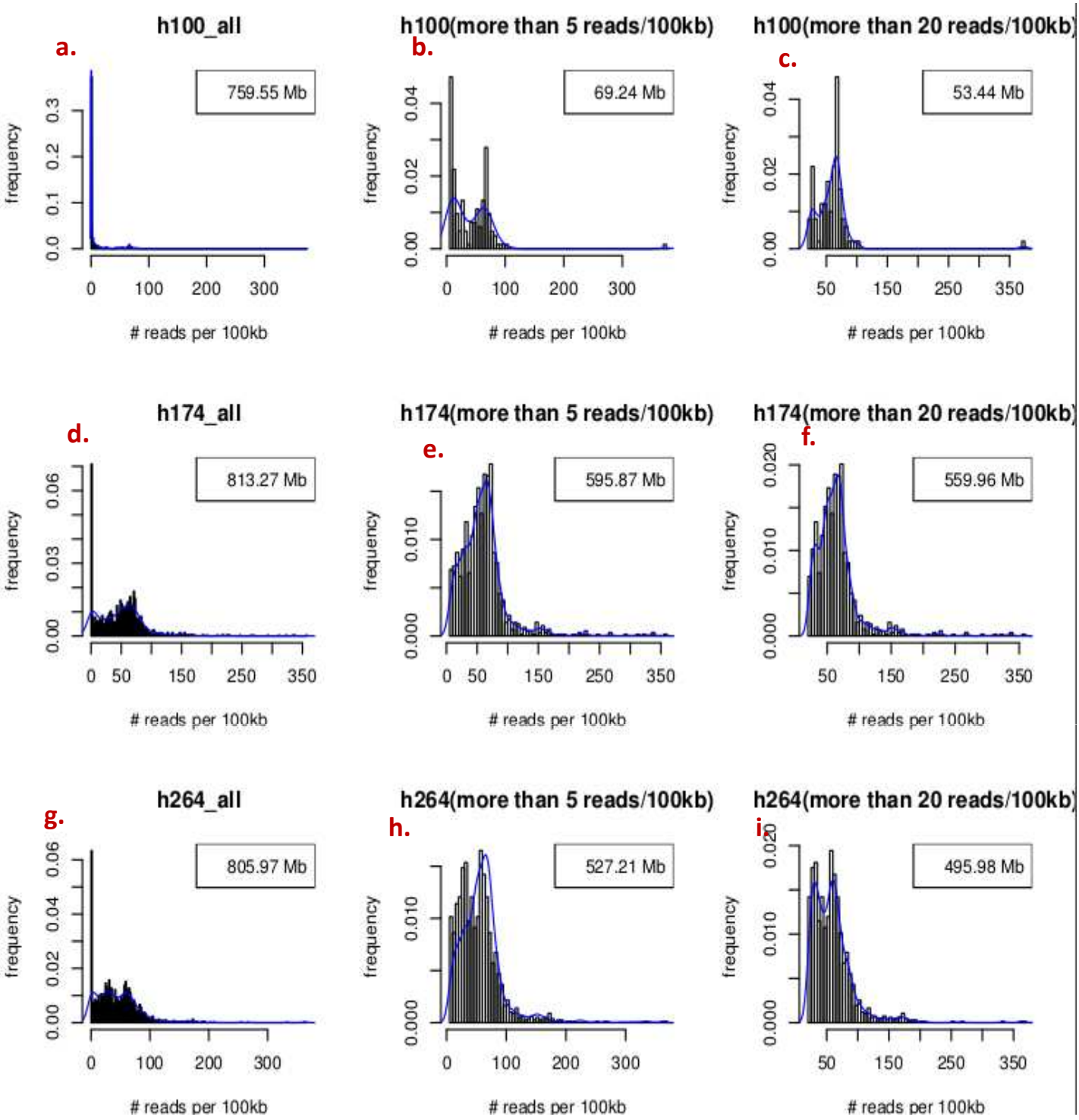


Figure1A: Read coverage for three different hybrid clones. The distribution of the number of reads per scaffold is corrected for the scaffold size (values are given in terms of # reads /100kb). The blue line depicts the empirical density. For each hyrbid, the three histogram represents, the read coverage, in #reads per 100kb, for (left) scaffold with at least one read, (center) scaffolds with at least 5 reads/100kb and (right) scaffold with at least 20 reads per 100kb. The value given within the box represent the total length of the involved scaffolds. This total length does not reflect the total length of duck fragments in the hybrids because some scaffolds can be broken.

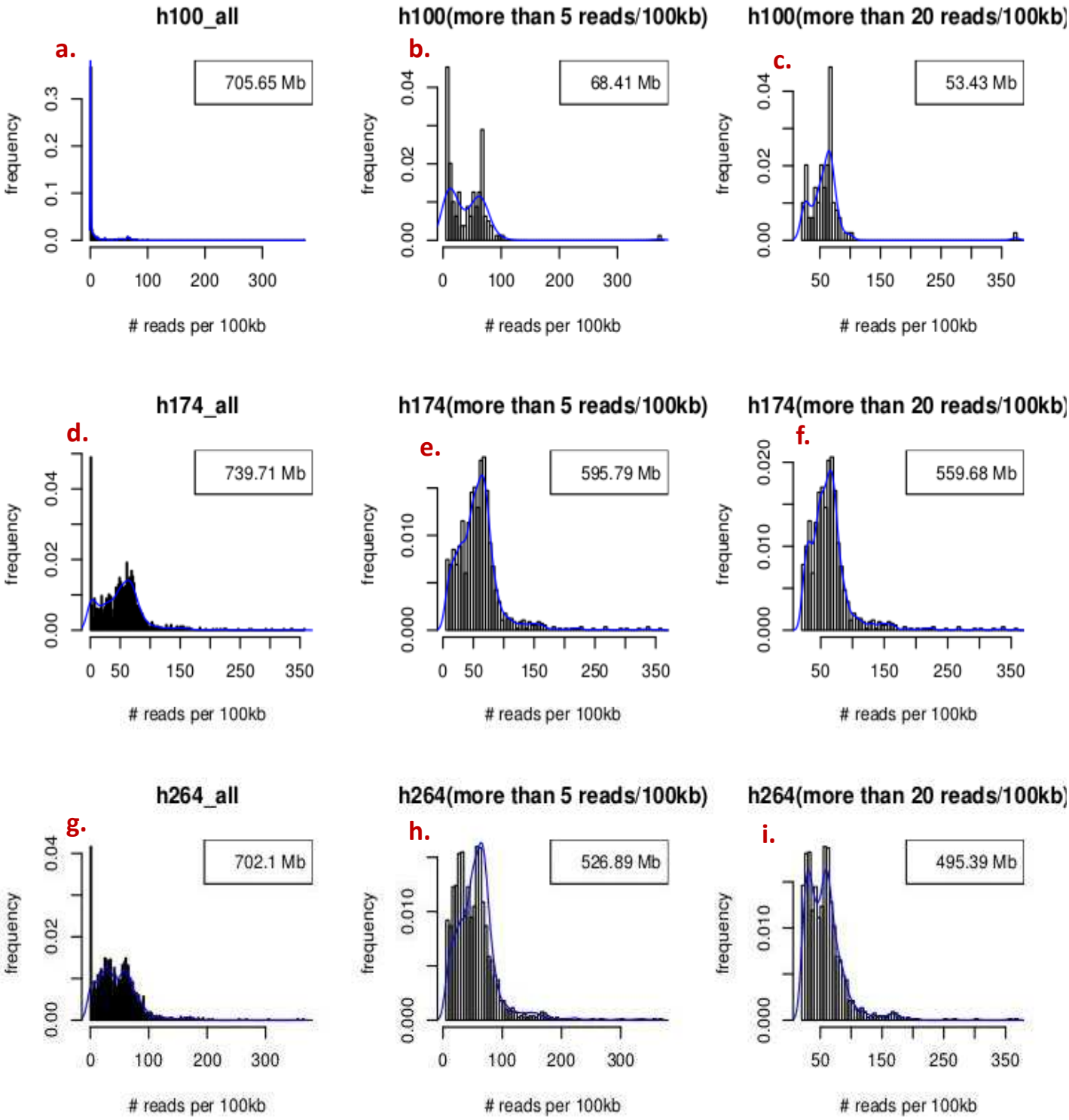


Figure1B: Distribution of read coverage (# reads /100kb) after removing reads mapped on hamster. The distribution of the number of reads per scaffold is corrected for the scaffold size (values are given in terms of # reads /100kb). The blue line depicts the empirical density. For each hyrbid, the three histogram represents, the read coverage, in #reads per 100kb, for (left) scaffold with at least one read, (center) scaffolds with at least 5 reads/100kb and (right) scaffold with at least 20 reads per 100kb. The value given within the box represent the taotal length of the involved scaffolds. This total length does not reflect the total length of duck fragments in the hybrids because some scaffolds can be broken. **Removing the reads that maps also to the hamster genome does not change the general picture.**

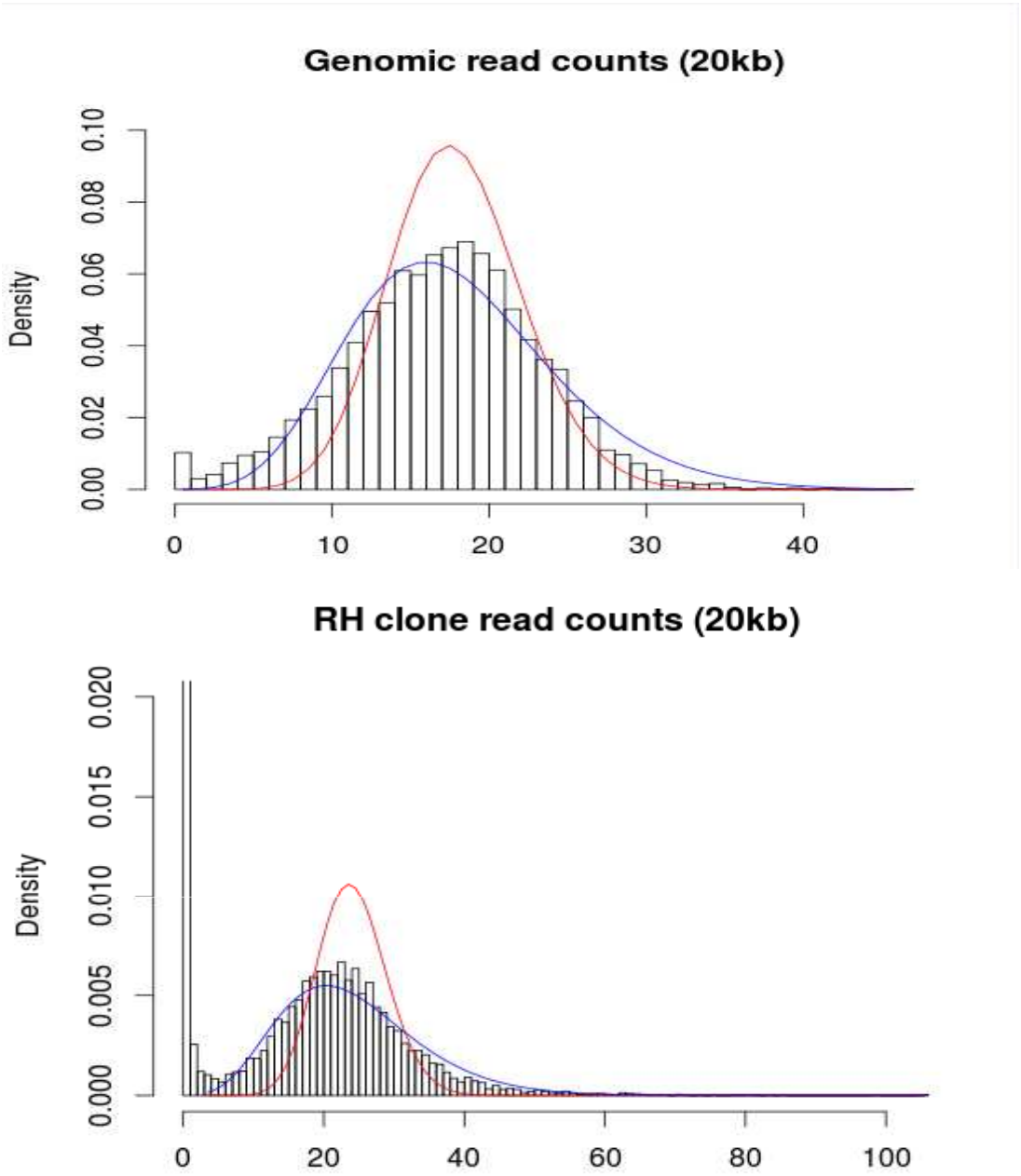


Figure2: Read coverage, expressed in terms of number of reads in 20kb windows. Top: read coverage distribution observed when sequencing genomic DNA (chicken genomic sequence reads kindly provided by F.Pitel). Bottom: read coverage distribution with sequencing reads obtained by sequencing the RH panel. Red and blue lines correspond respectively to a fit with the Poisson and negative binomial distribution. The well known over dispersion of the distribution of read counts can be observed here with the departure to the Poisson dispersion. This over dispersion is even more pronounced with the sequencing reads originating from the RH panel sequences. This more pronounced over dispersion suggests that the variation in read coverage is not only the result of the sequencing bias also observed with genomic data, but also reflects the fact that the hybrids are a mixture of cells with different genomic content.

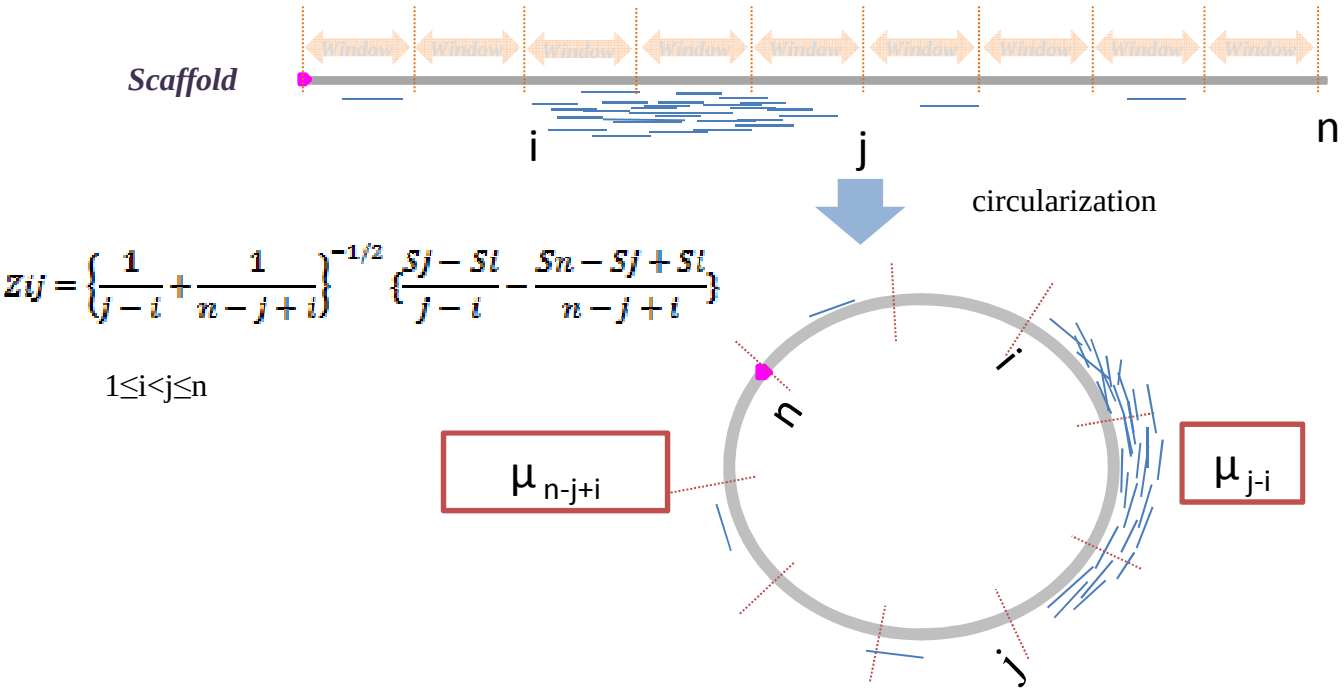


Figure 3: **principle of circular binary segmentation (CBS)**

CBS first circularize the scaffold (start indicated in pink). A sliding window size (e.g. 20kb) is fixed and reads are counted for all non-overlapping windows providing n ordered observation X_1, …, X_n. We define S_i as the partial sum ΣX_j .The CBS searches recursively for segments of different means using the statistics $Z_C=\max_{1\leq i\leq j\ \leq n}|Z_{ij}|$. The significance of the statistical test is judged by permutation. The binary circular segmentation procedure applies the test recursively until no changes are detected in any of the segments obtained from the change-points already found.

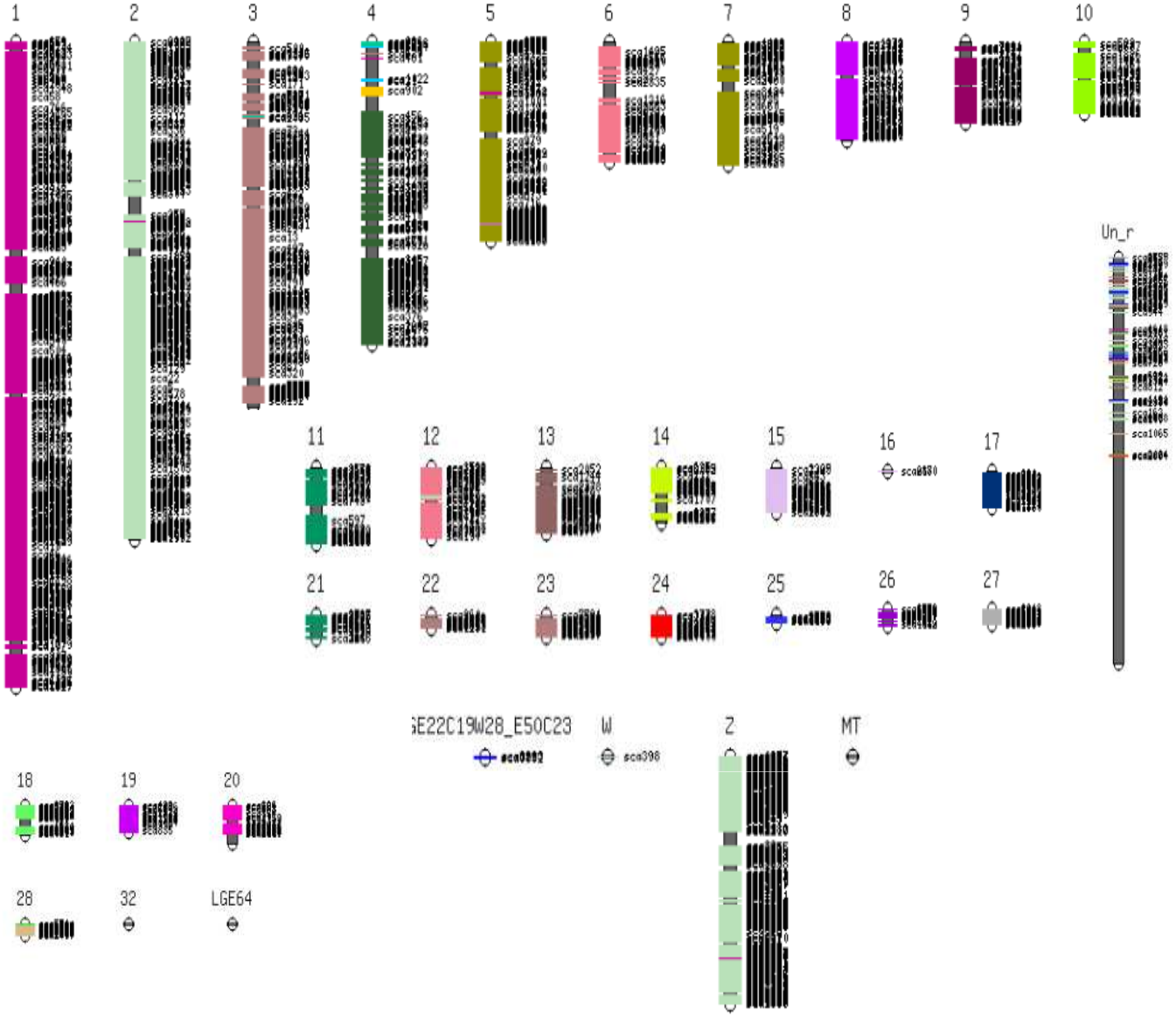


Figure 4: Fifty-one RH linkage superimposed on chicken chromosomes.
All RH vectors were subjected to RH linkage analysis, from which 51 RH linkages were obtained by LOD threshold of 4.5. Then these 51 RH linkages were superimposed on chicken chromosomes as shown in the figure. Each color was a linkage group. The markers having extremely high retention were filtered for analysis, thus the GGA4p was not well covered. The linkage analysis could also assign some scaffolds which were mapped to chicken ChrUN to a specific chromosome.

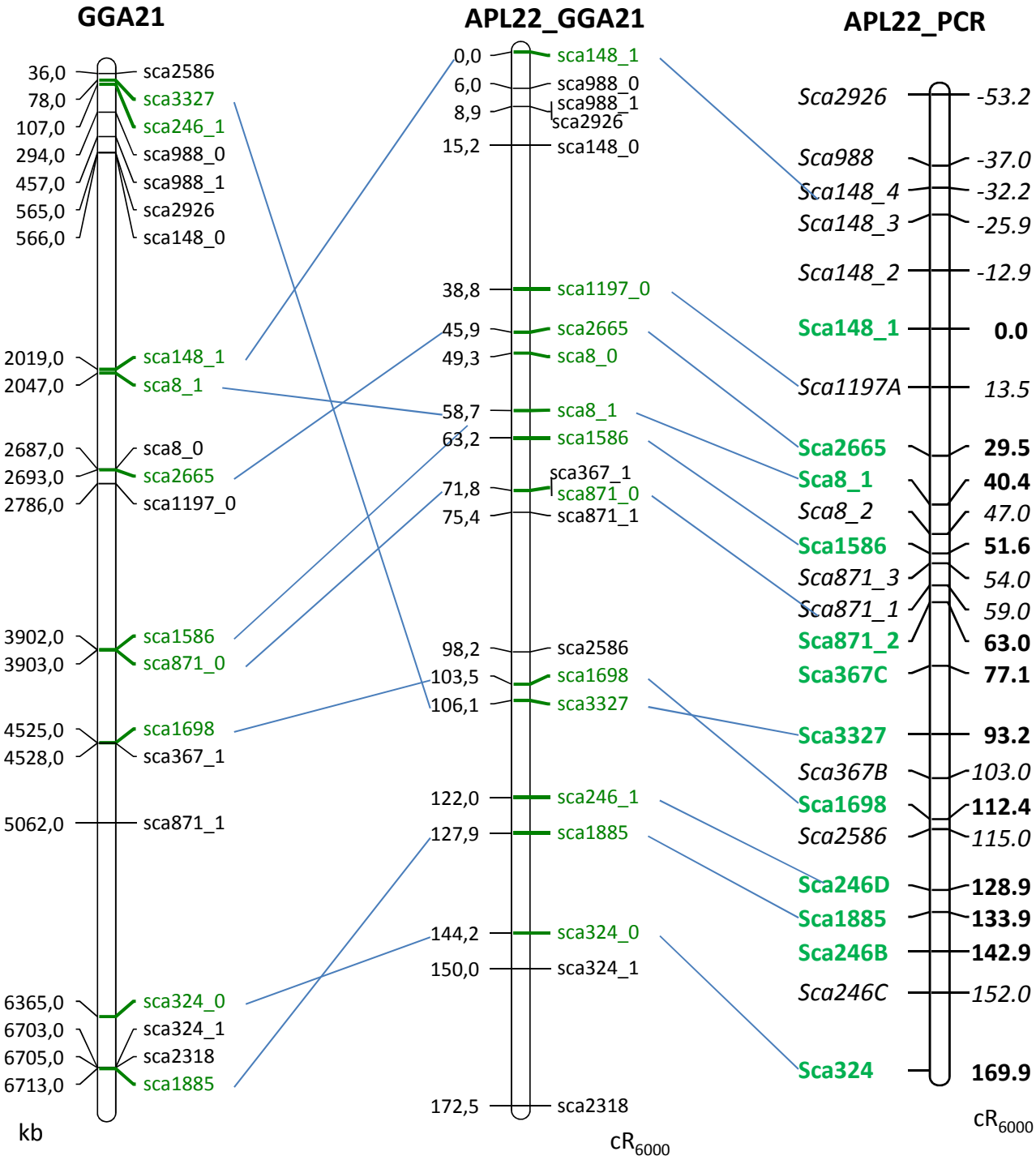


Figure 5: Comparison of APL22 RH maps obtained by sequencing or by PCR. Left: scaffold position on GGA21, middle: RH map of APL22 built by the comparative mapping method and based on the data from sequencing the hybrids. Right: RH map of APL22 based on conventional genotyping on the whole genome (WGA) amplified panel (Rao et al., accepted). The nomenclature of the markers in the maps on the left and middle is different from the map on the right. Markers in the maps on the left and right have suffixes "_0" and "_1", used for orientating scaffold: "_0" is the beginning of the scaffolds and "_1" is the end of scaffold. Markers on the right are PCR amplicons designed from scaffold sequence, whose accession numbers are given in Rao et al, accepted. Markers in green are robust or framework map markers whose position are quite certain. Only robust map markers on APL22_GGA21 are linked by blue lines.

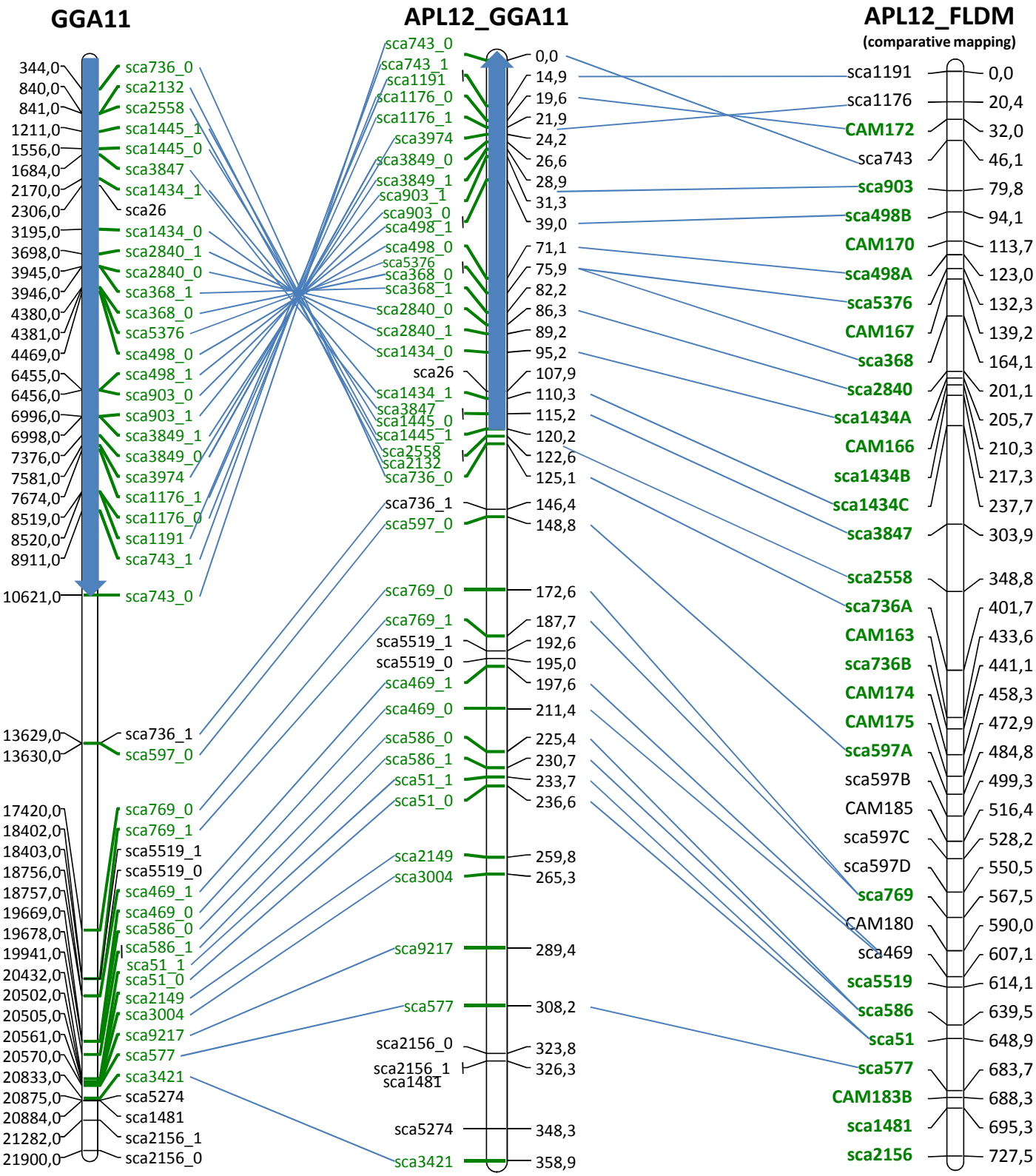


Figure 6: Comparison of APL12 RH maps obtained by sequencing or by PCR. Left: scaffold position on GGA11, middle: RH map of APL12 built by the comparative mapping method and based on the data from sequencing the hybrids. Right: RH map of APL12 based on FLDM genotyping (Rao et al., accepted). The nomenclature of the markers follows that of figure 5.

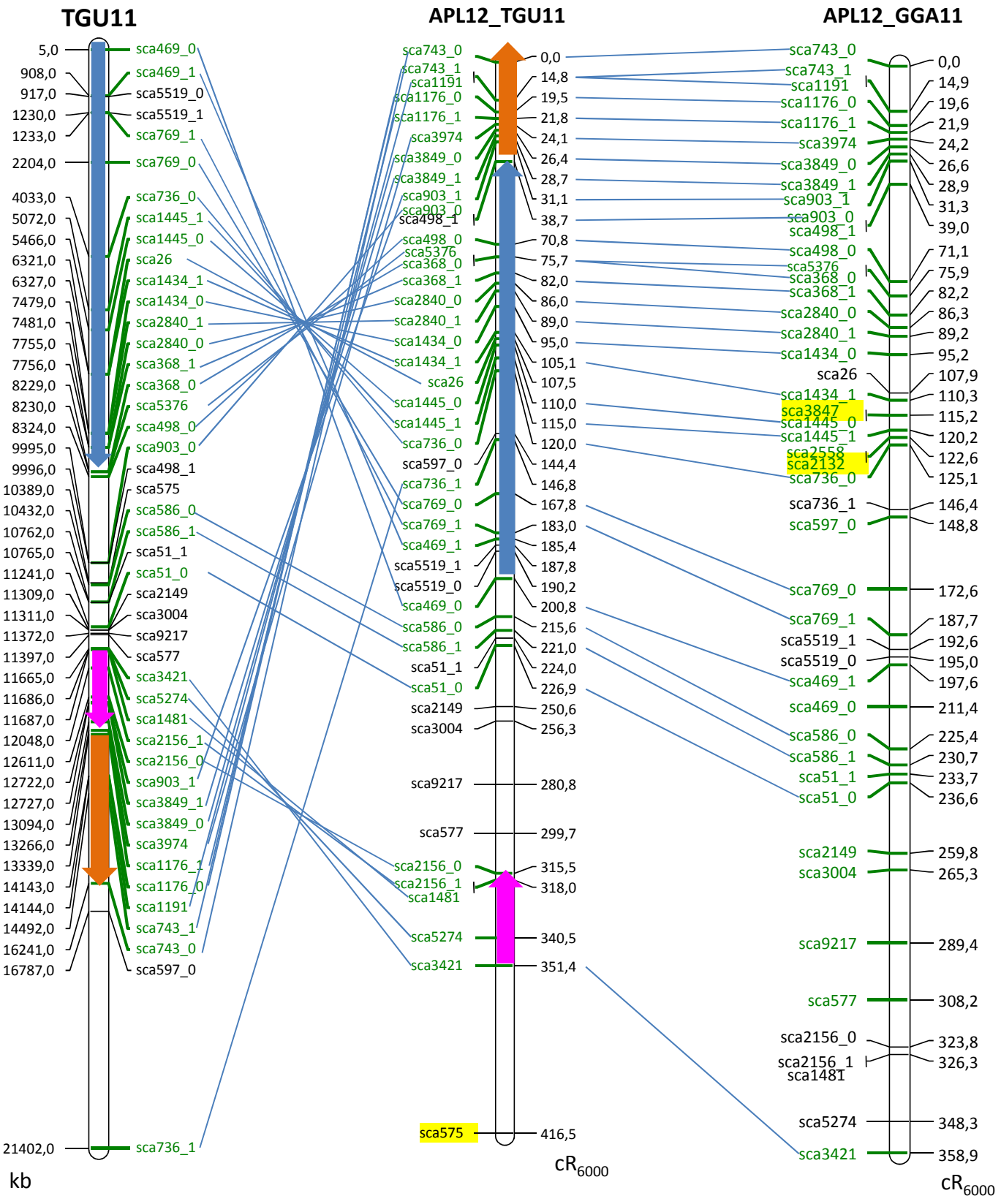


Figure 7: Comparing RH maps obtained with chicken or zebrafinch as reference. Left: markers placed on TGU11, middle: RH map of APL12 made by comparative mapping using zebra finch as reference. Right: RH map of APL12 made by comparative mapping using chicken as reference. Nomenclature of markers are as in figure 5. Makers in green were on the robust map. Markers highlighted in yellow are not on both maps due to non-alignment on one of the reference genomes. Only robust map markers are linked by blue lines. Arrows indicated inversion.

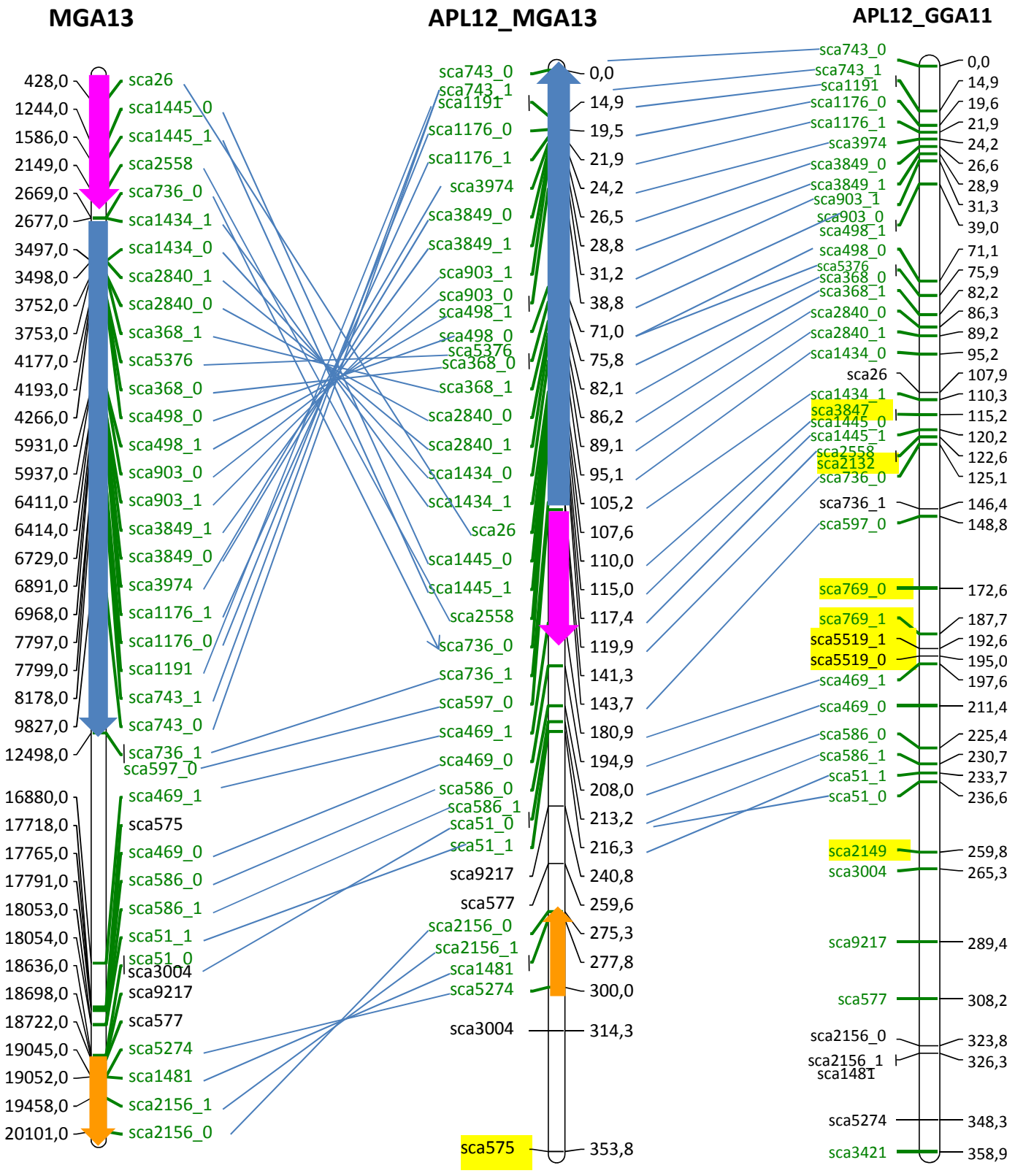


Figure 8: Comparing RH maps obtained with chicken or turkey as reference. Left: markers placed on MGA13, middle: RH map of APL12 made by comparative mapping using turkey as reference. Right: : RH map of APL12 made by comparative mapping using chicken as reference. Nomenclature of markers are as in figure 5. Makers in green were on the robust map. Markers highlighted in yellow are not on both maps due to non-alignment on one of the reference genomes. Only robust map markers are linked by blue lines. Arrows indicated inversion

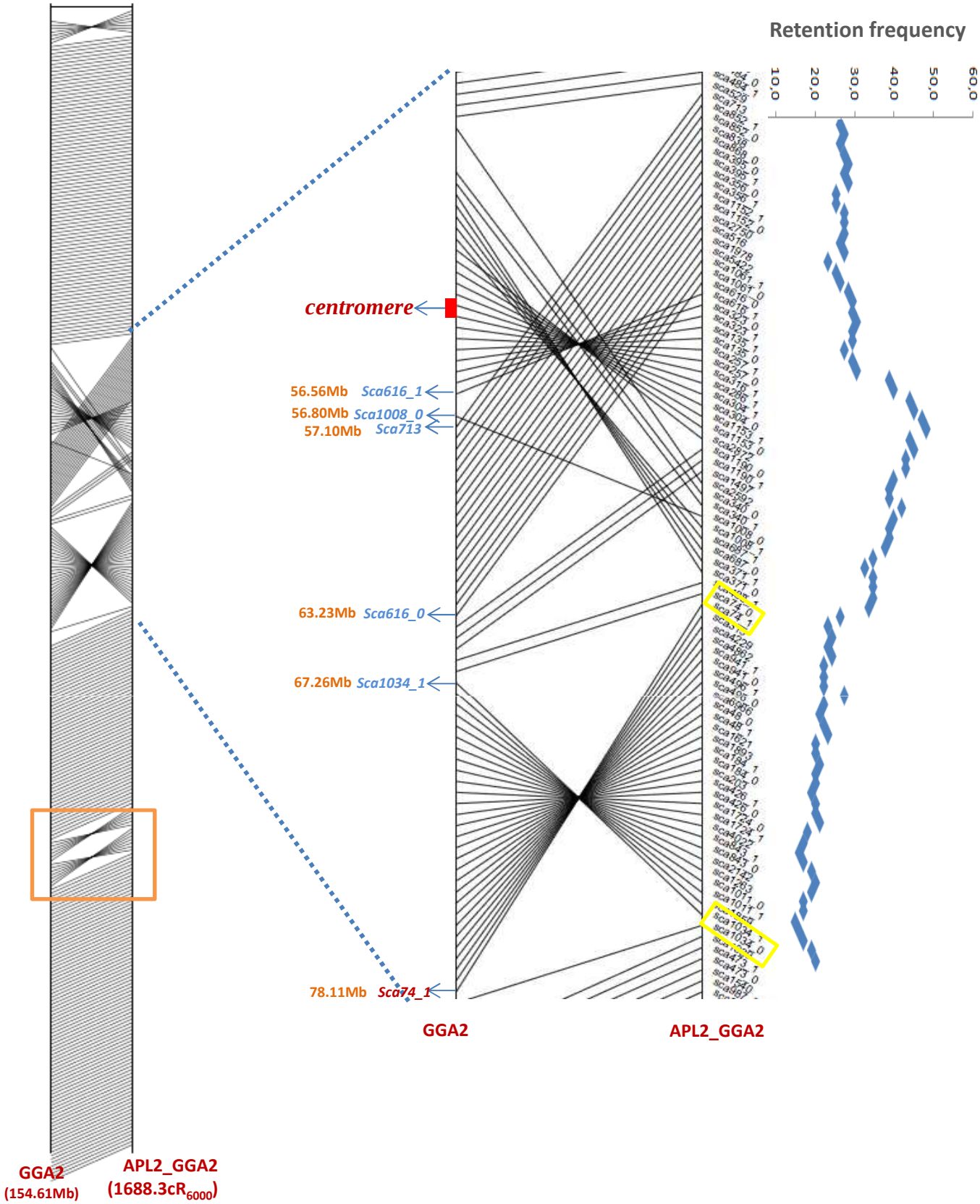


Figure 9: APL2-GGA2 comparative map. Left: comparative map of the APL2 robust map with the GGA2 sequence. Middle: zoom-in of the major rearrangements region. Position (Red) and scaffold name (blue) of six chicken WAG BAC clones used for FISH mapping. A red bar indicates the centromere position in chicken. The yellow box showed indicated scaffolds in which the breakpoints take place. Right: the retention of the markers located in the zoomed-in region. The orange box shows the two inversions specific to the duck genome.

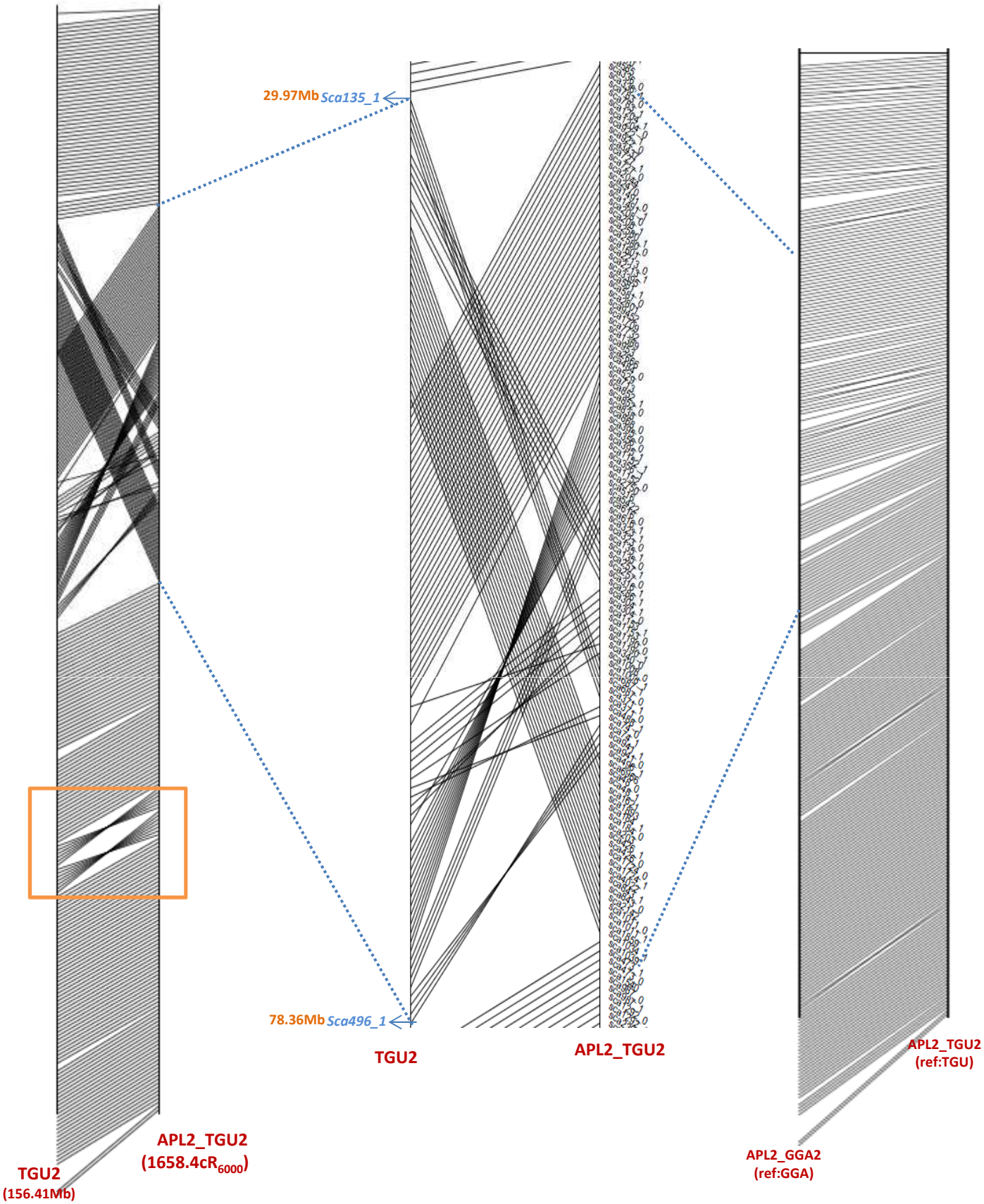


Figure 10: APL2-TGU2 comparative map. Left: comparative map of APL2 robust map with the TGU2 sequence. Middle: zoom-in view of the major rearrangement region. The name and coordinates on zebra finch markers located on the border of the rearrangements are indicated. Right: comparison of robust maps of APL2 obtained by using GGA2 or TGU2 as reference genomes, showing the consistency in the results. The orange box shows the two inversions specific to the duck genome.

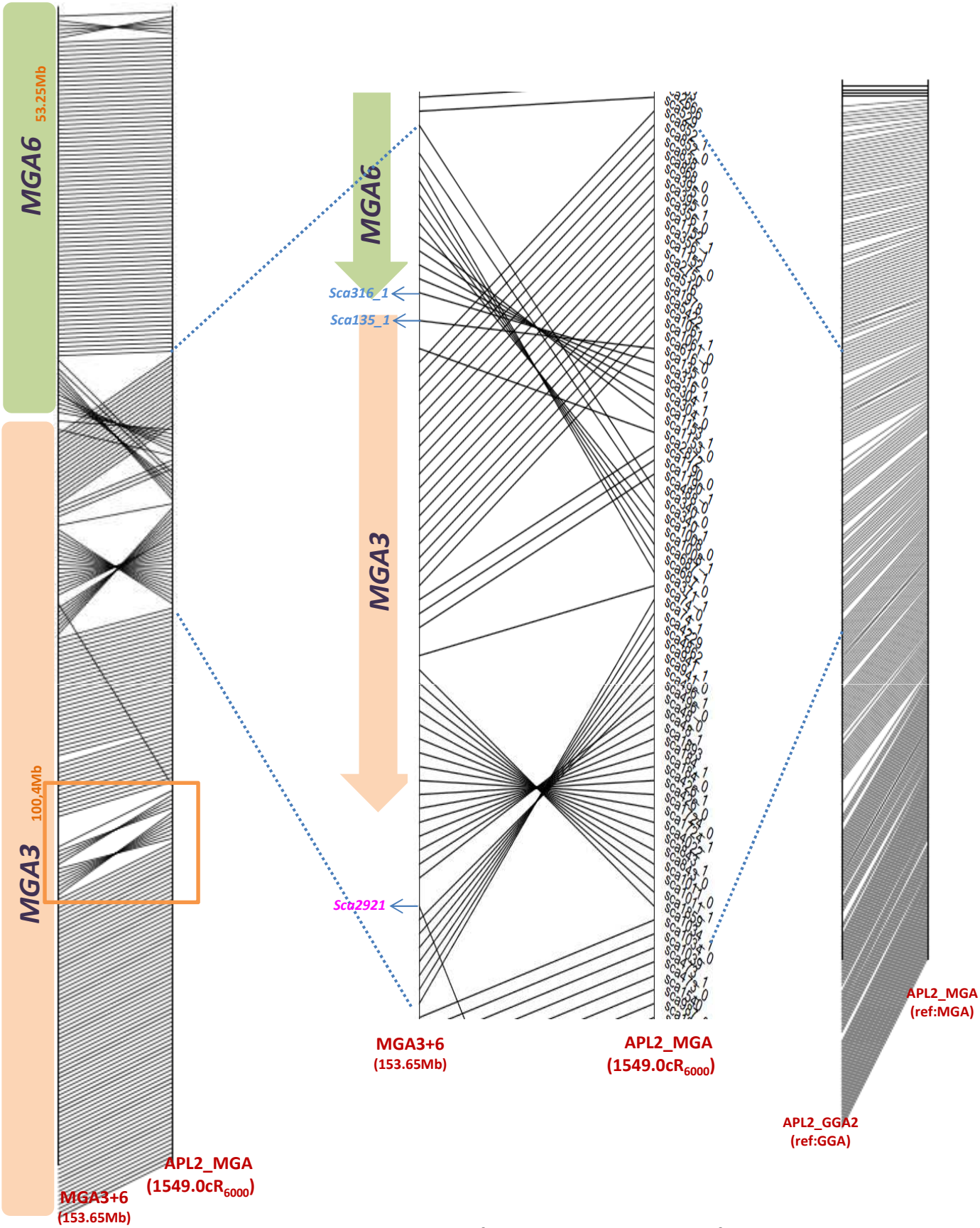


Figure 11: APL2-TGU comparative map. Left: comparative map of APL2 robust map with MGA6+MGA3. Middle: zoom-in view for the major rearrangement. The scaffolds in the MGA6+MGA3 junction point are indicated in blue. The single transposed scaffold is indicated in pink. Right: comparison of robust maps of APL2 obtained by using GGA2 or MGA(6+3) as reference genomes, showing the consistency in the results. The orange box shows the two inversions specific to the duck genome.

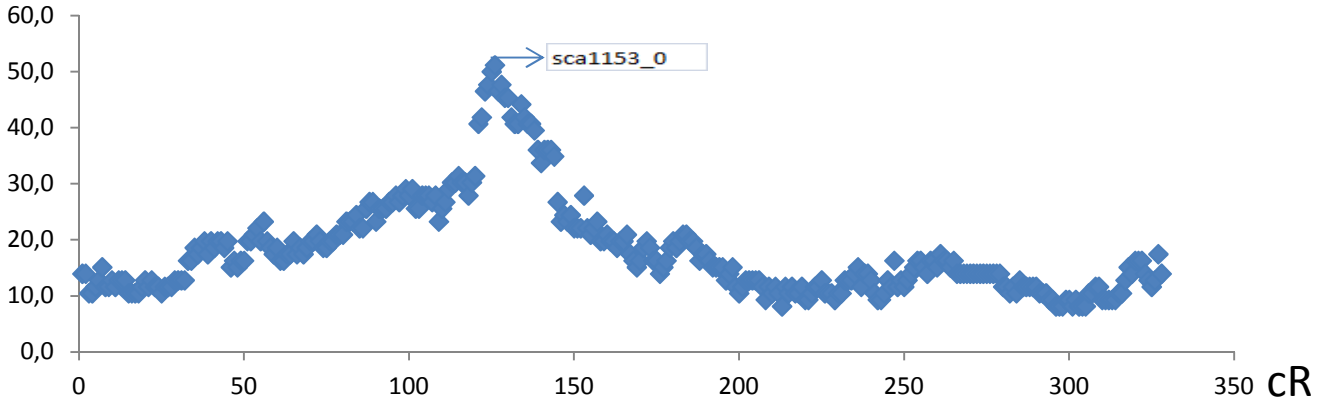


Figure 12: retention frequency of APL12 RH markers. RH map of APL12 built using using chicken as a reference. The highest retention is for sca1153_0, with 51.2%, suggesting the centromere position. The average retention for this chromosome is 18.2%.

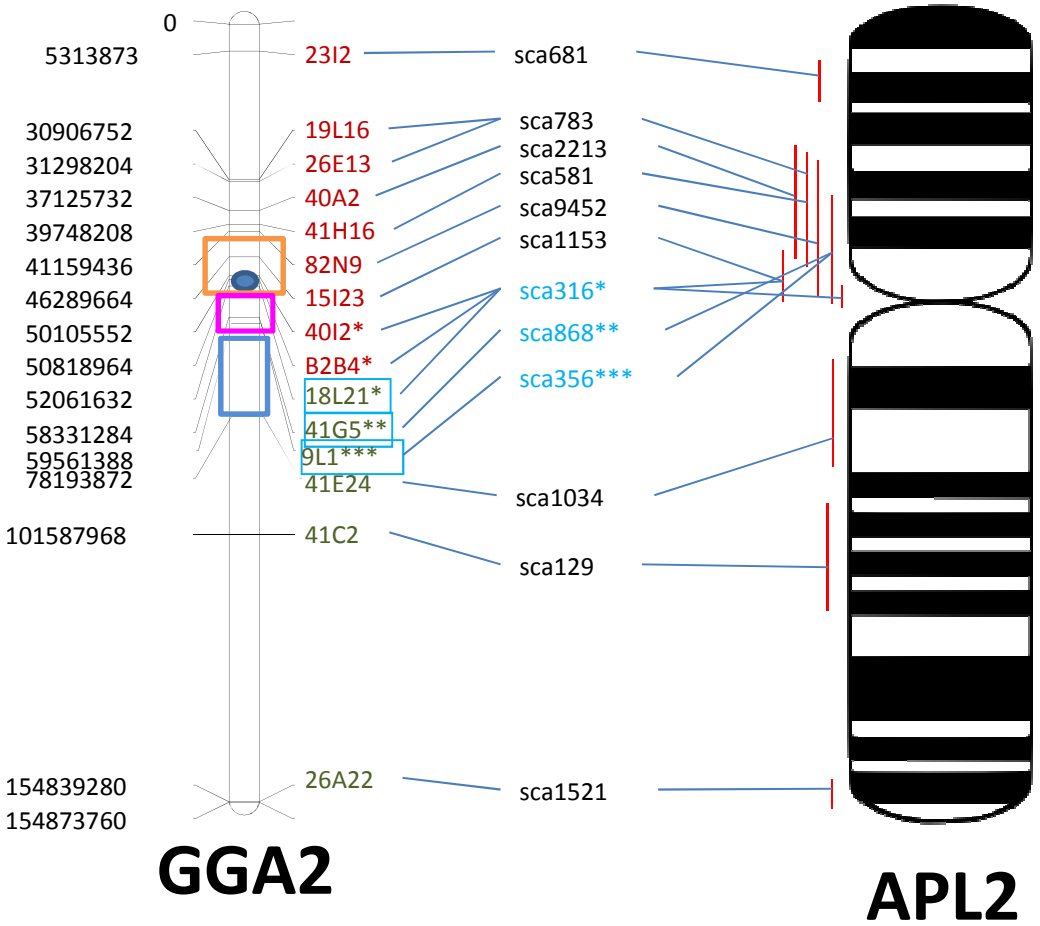


Figure 13: APL2-GGA2 rearrangements, as shown by FISH data in the litterature. FISH mapping data of chicken BAC clones hybridised to duck metaphase chromosomes, provided by Fillon *et al* (2007) and Skinner *et al* (2009). Chicken BAC clones were hybridized on duck metaphases. Red bars indicate the position of the signal on the duck chromosome and the blue line inferred the best position. Sca316 had three corresponding BAC clones distinguish by a single "*" and spanning the chicken centromere: two on GGA2p and one on GGA2q. Sca868 and sca356 had a single BAC for each notified by"**" and "***" respectively. The FISH data demonstrated that they were located on the GGA2q but in the short arm in duck. The orange box indicates the position of the inversion between chicken and duck suggested by RH maping. The pink boxe the region where a translocation is detected. The blue box suggest the position of the 10Mb inversion on GGA2q.

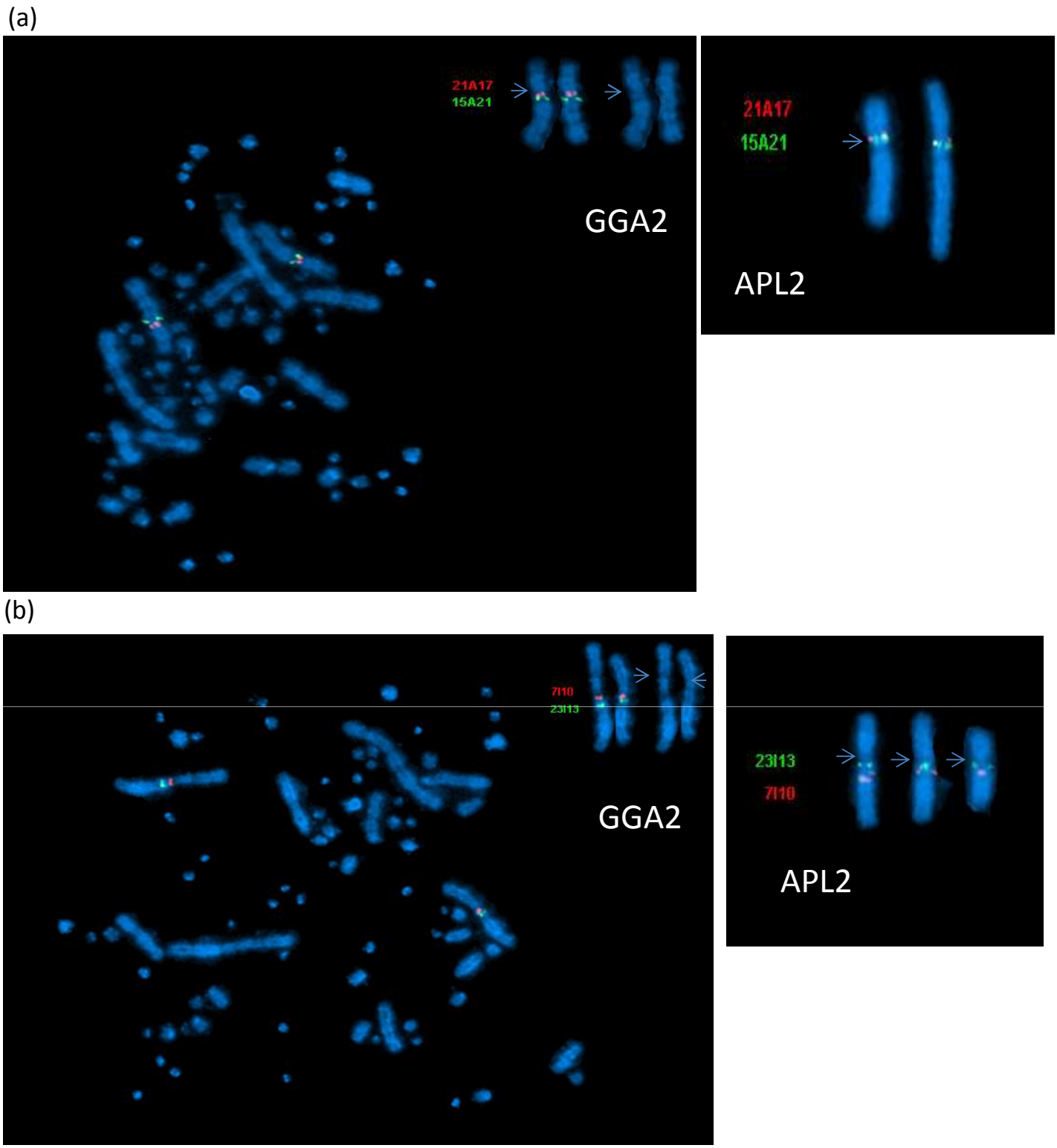


Figure 14: Confirmation of GGA2-APL2 rearrangement by FISH. (a) confirmed translocation from region sca713 to sca616_0 on the map. BAC clone 21A17 corresponded to sca713 in red while 15A21 corresponded to sca616_0 (in green). In GGA2 the signals were on q arm whereas in duck were on p arm. As the region was near centromere, the chromatin was condensed thus the interval seemed small on chromosome. (b) Confirmed inversion spanning sca1034_1 to sca74_1 on GGA2q. BAC clone 23I13 corresponds to sca74_1 while 7I10 corresponds to sca1034_1. This inversion was about 11Mb. Centromere positions are indicated by arrows.

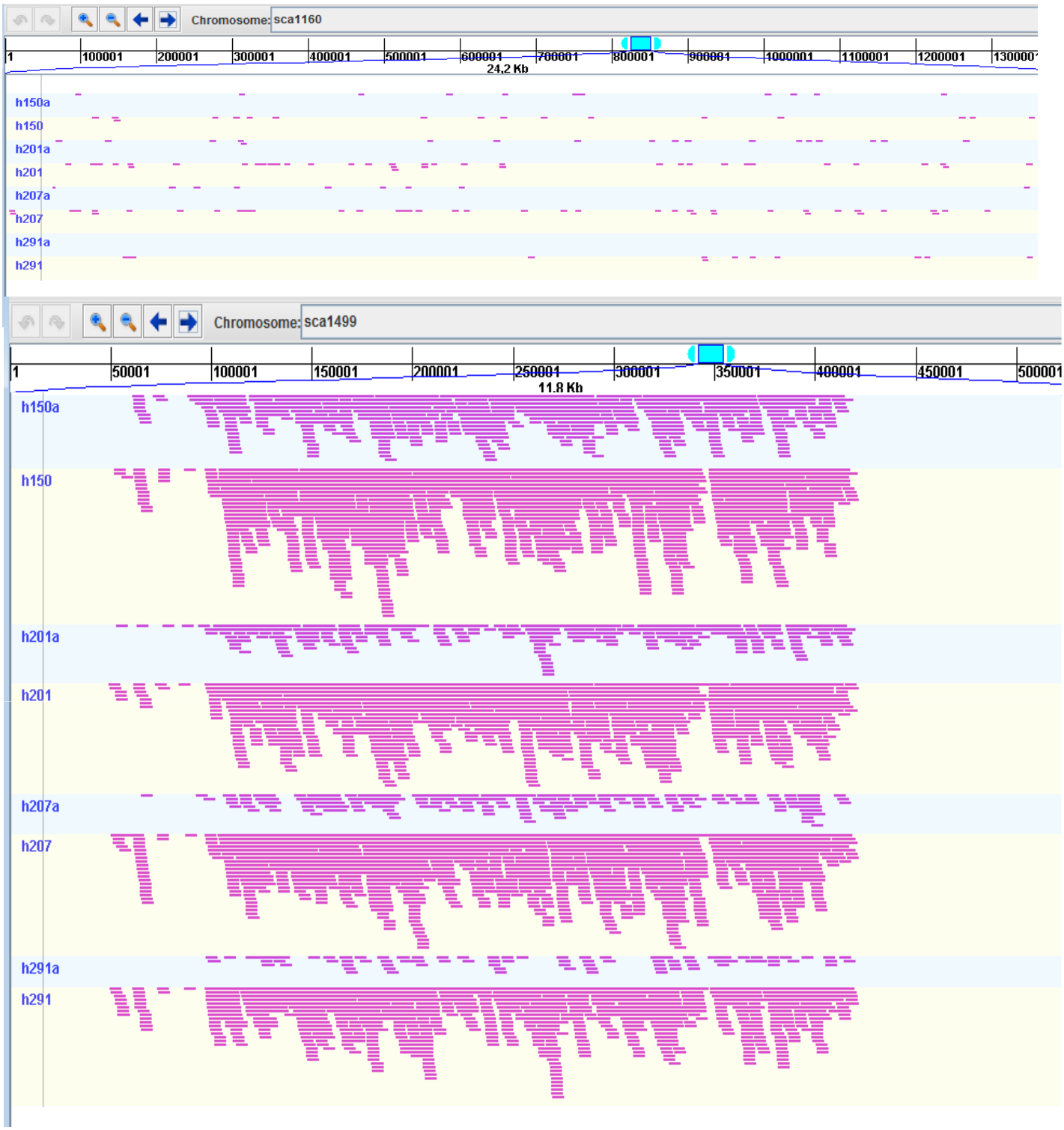


Figure IV-1: Sequencing reads on WGA hybrids and *ATG4A* copy number. Visualization of sequence alignments with GenomeView for 4 hybrids, either WGA amplified: e.g. h291a or not: e.g. h291. Top: sca1160 is the scaffold bearing *HPRT* and therefore is retained in all the hybrids. The zoom is on the *HPRT* gene. Bottom: sca1499 bearing the *ATG4A* gene. The zoom is on *ATG4A*. In all the hybrids shown above, it is clear that *ATG4A* had much higher read coverage than *HPRT*. Amplified hybrids have a lower read coverage, including for *HPRT*, the selection gene.

Complementary results and discussion

A highly repeated gene in duck genome: *ATG4A*

Before performing the segmentation process and genotype calling of the scaffolds, we investigated the scaffolds that had extremely high retention. Not surprisingly, all the scaffolds were located near the *HPRT* gene used for the selection of hybrids containing duck chromosome fragments. A notable exception was sca1499 which is be located on APL2 according to comparative mapping data with chicken. Furthermore, we also tried to explore whether there was a link between the proximity to *HPRT* gene and read coverage. Therefore we visualized those data with GenomeView program (Figure IV-1) and interestingly, we found out that the sca1499 had distinct pattern in term of read coverage. The *HPRT* gene is located in sca1160, which is evenly covered in most of the hybrids, whereas sca1499 presents a very high read coverage only for a region of about 8 kb in most of the hybrids. Moreover, the rest of sca1499 has a significantly low read coverage, when the whole scaffold is present. We searched for this 8 kb fragments in the human and chicken genomes by BLASTN and found a high sequence similarity to the *ATG4A* gene. This gene involves in the process of autophagy which is a major catabolic pathway by which eukaryotic cells degrade and recycle macromolecules and organelle (Scherz-Shouval et al. 2007). *ATG4A* is one of four homologs of *ATG4* which cleave *ATG8* and thereby allow the conjugation and deconjugation of cleaved *ATG8* and phosphatidylethanolamine (PE). The conjugated ATG8-PE functions in membrane dynamics during autophagy (Scherz-Shouval et al. 2003).

The most intriguing finding was that this gene is present in its integrality at a high copy number in all the hybrids. We first speculated that multiple copies might result from gene amplification since previous studies showed that some genes can amplify under the selection stress in somatic hybrids (Schimke 1984; Stark 1986). As the sequencing data showed a very high read coverage, reflecting a very high copy number, we tested whether

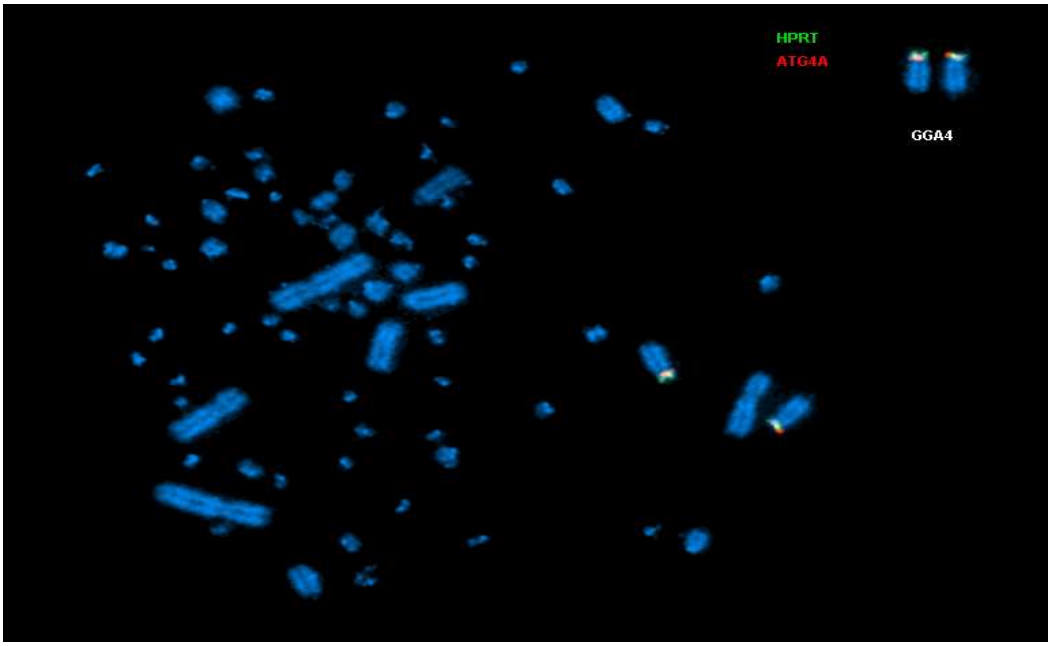


Figure IV-2a: localization of *HPRT* and *ATG4A* in chicken. The FISH results are in good agreement with the sequence data. The *HPRT* gene locates at 4Mb on GGA4p whereas *ATG4A* at 14Mb in GGA4p.

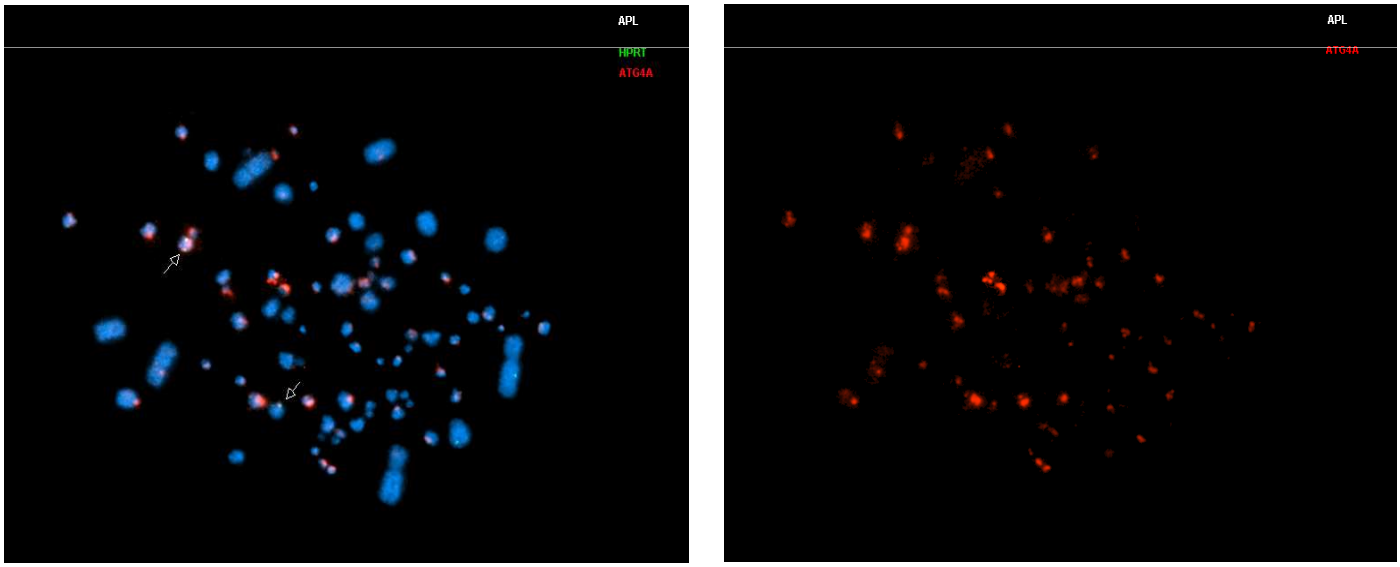


Figure IV-2b: localization of *HPRT* and *ATG4A* in duck. The *HPRT* gene locates on a microchrosomome, consistent with previous cytogenetic studies (Skinner *et al*, 2009), whereas the *ATG4A* gene is highly repeated in the duck genome. Right: the same metaphase as on the left, only the signal of *ATG4A* gene is shown. Microchromosomes show higher intensity than macrochromosomes.

those units were arrayed in tandem or dispersed. In addition, we also tested whether this gene was co-retained with the *HPRT* gene. In this context, we performed FISH experiment on chicken, duck and duck hybrids metaphases (FigureIV-2). Strikingly, the results show that the *ATG4A* gene does not appear as specifically amplified in the hybrids, but to be present in the duck genome at a high copy number and dispersed throughout the genome on most chromosomes, with a preferential enrichment on microchromosomes. Moreover, the FISH experiments showed that *ATG4A* is present as a single copy un the chicken genome (FigureIV-2a,b).

In human, there are 4 homologous genes to *ATG4A*, in which *ATG4B* has the broadest substrate spectrum with similar affinity and catalytic efficiency toward each of ATG8 substrates and then followed by *ATG4A* (Li et al. 2011a). We have found that duck has an ortholog of the *ATG4B* gene in sca2210; however, the sequencing data does not suggest that *ATG4B* is amplified or has multiple copies.

The *ATG4A* gene is 8kb long and apparently expanded in the duck genome. Therefore it is not clear why this gene was not detected as high copy in the genome sequencing and assembly. The only assembled copy of *ATG4A* is on a scaffold belonging to APL2. Although the FISH mapping confirms the presence of the gene on APL10, close to *HPRT*, this copy does not appear to have been assembled. *ATG4A* gene has twice the length of full length CR1 repeats (Wicker et al. 2005). If this gene is a novel repeat which appeared in the duck genome, its GC content is close to the genome average and larger than that of any known LINE family(Mathias et al. 1991; Wicker et al. 2005). Moreover, it seems enriched on microchromosomes, as opposed to CR1 repeats which are enriched on macrochromosomes.

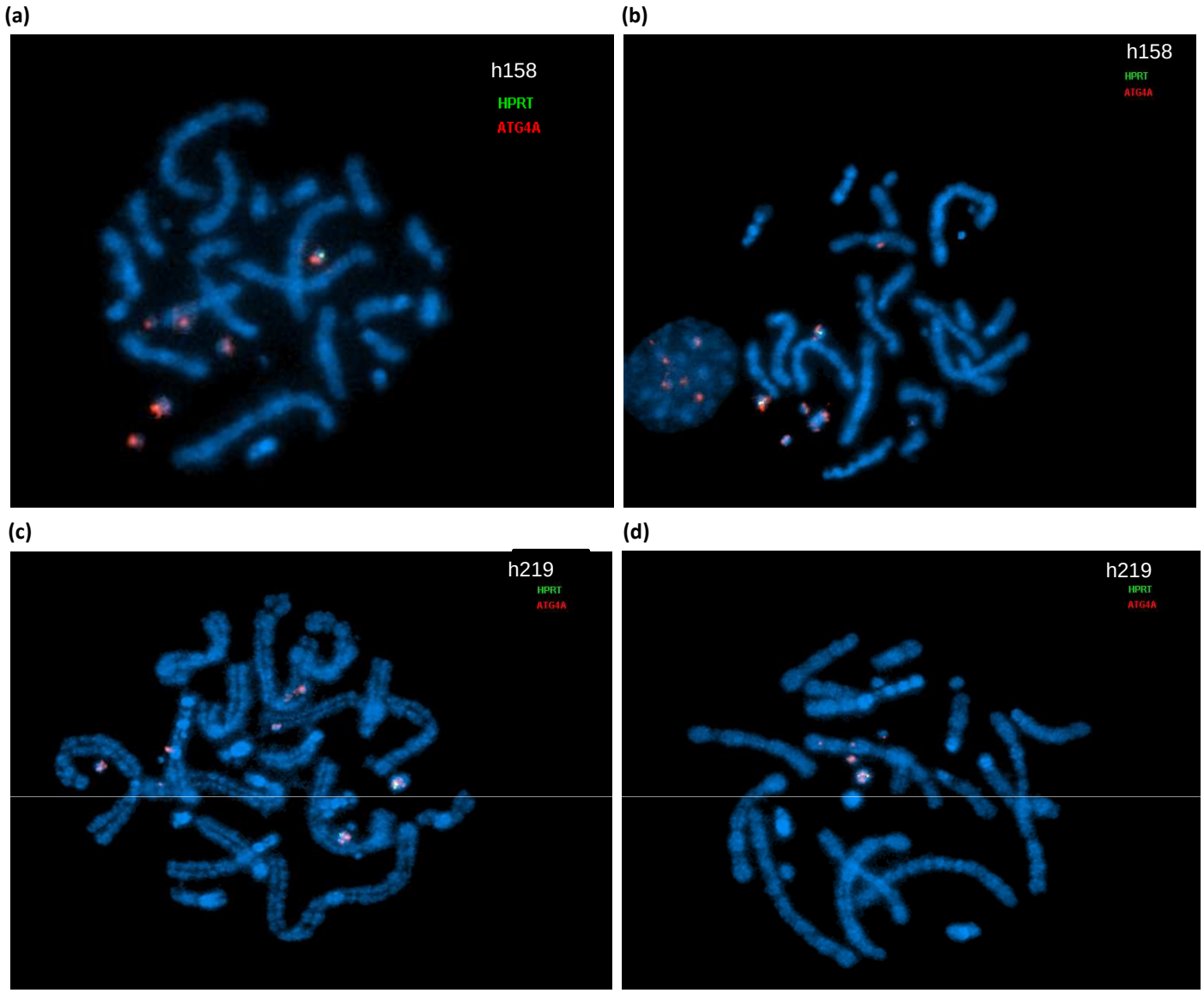


Figure IV-2c: Localization of *HPRT* and *ATG4A* in duck hybrids. Two hybrids were chosen: h158 for (a) and (b) ; h219 for (c) and (d). For each hybrid, two metaphases are shown. Combining the results on duck metaphase, these results above reflected that *ATG4A* and *HPRT* are not associated. (a) and (b) showed no *ATG4A* signal in hamster chromosomes in hybrid h158, (c) was the same as in hybrid h158. whereas (d) showed the *ATG4A* gene inserted into hamster genome, suggesting that a few duck fragments could insert in the hamster genome, which had not been detected by hybridizing duck genomic DNA on the hybrids.

Sequencing whole genome amplified (WGA) hybrids

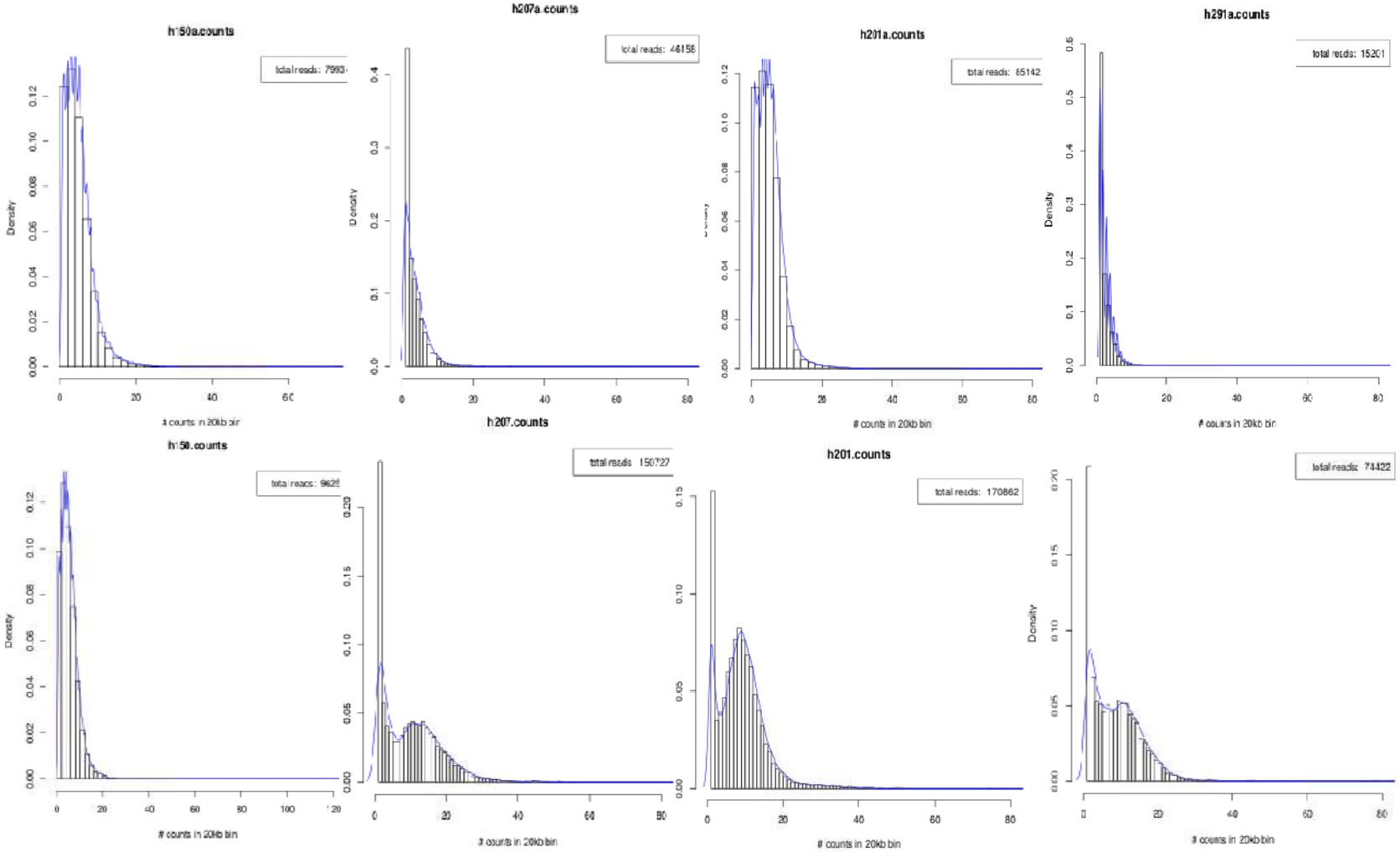
The duck whole genome RH panel has not been cultured at a large scale. Instead, we used whole genome amplification (WGA) with the Multiple Displacement Amplification (MDA) method to obtain large quantity of DNA (chapter II). It was previously reported that the WGA by MDA method allowed a relatively unbiased amplification and was subsequently applied in single cell sequencing for cancer research (Hou et al. 2012; Xu et al. 2012). However, the genotyping of 8 nohit markers (chapter II) revealed also that some genomic regions were difficult to amplify (Rao et al. 2012), so we chose 4 WGA hybrids to sequence and compare to the non-WGA sequence data, to see if WGA hybrids are suitable for the genotyping by sequencing method. The primary sequencing output is shown in Table IV-1, from which it is clear that WGA hybrids have less reads mapped to duck than non-WGA hybrids. Moreover, read alignment to scaffolds also showed that results of scaffolds genotype calling was not always consistent between WGA and non WGA hybrids (data not shown). We visualized the alignment of sca1160 in the *HPRT* region and sca1499 in the *ATG4A* region with GenomeView (Figure IV-1). The WGA hybrids have a similar trend in sequencing depth variation around *ATG4A,* than the non WGA hybrids, although with a lower read coverage. Although the *HPRT* gene, as a selection gene for donor cell chromosomes, is retained in all four hybrids, the sequencing depth is lower in the WGA hybrids. Moreover, in one case, no read was observed in the amplified hybrid, despite the gene being present.

Since the low read coverage data in the WGA hybrids caused uncertainties in the calling process, we hence compared the read coverage distribution for each hybrid (data shown in FigureIV-3). From FigureIV-3 it is clear that the WGA hybrids contain a larger proportion of scaffolds with low read coverage. Non WGA hybrids gave better data and thus the WGA hybrids were discarded for further analysis. In addition, as WGA DNA was difficult to quantify (Rao et al. 2012) and as the DNA concentration is an important parameter in NGS, we chose non WGA hybrids for sequencing. Although we have shown that WGA DNA was

HYBRID	Nb of Reads produced	Nb of Reads pair mapped on duck
h150a	**4,773,538**	**94,806**
h150	**4,346,000**	**130,199**
h201a	**4,393,723**	**98,172**
h201a	**6,656,681**	**220,509**
h207a	**3,116,217**	**52,205**
h207	**6,793,978**	**199,571**
h291a	**1,647,997**	**18,041**
h291	**5,276,147**	**97,090**

TableIV-I: primary sequencing and alignment on duck of WGA and non-WGA hybrids. Reads mapping both to duck and hamster were not removed. Nomenclature for normal hybrids is: h+arabic number whereas whole genome amplified hybrids had an “a” as suffix.

difficult to quantify which could probably be the cause for the lower number of sequencing reads such as in h291a, h150a showed a higher number of raw sequencing reads than h150, suggesting that in the WGA hybrids, the lower number of mapped reads might be due to a lower proportion of usable reads, when compared to non WGA hybrids.



FigureIV-3: distribution of reads counts per 20 kb bin for read-containing scaffolds in WGA and non-WGA hybrids. Four hybrids are plotted, either WGA (top) or non-WGA (bottom). The total number of reads is indicated on the topright box in each plot graph. Reads mapping also to duck were removed only from the non-WGA hybrids. Despite this, it is clear that non WGA hybrids have a higher read coverage in the scaffolds.

Chapter V.

General Discussion and perspectives

The aim of the work presented here was to develop a mapping resource, the duck RH panel, and methods do generate rapidly genome-wide maps. The ultimate goal is to provide an improved duck genome assembly by ordering scaffolds along chromosomes. The whole genome RH panel is the basic infrastructure of the whole RH mapping system, so care must be taken in its construction. Therefore, we tested 4 different conditions to obtain duck radiation hybrids, from which one optimized protocol was selected for the RH panel construction. We suggest this protocol may be also adapted for other birds. We have then carried out two more fusion experiments using the optimized method to obtain a sufficient number of hybrids to select from for an optimized panel. A total of 225 hybrids were obtained, from which the 90 best ones were selected in the final panel. To avoid large scale cell culture, we compared three different genotyping techniques as described in Chapter III, and tested their efficiency by making two RH maps, illustrating the potential of this panel in improving duck genome assembly. From there, whole genome RH maps can be constructed. However, as a way of reducing the time and effort spent towards whole genome maps, we decided to sequence the whole RH panel and then align the sequencing reads to the duck scaffolds which are thus considered as the markers in this new approach. With the three maps presented in Chapter IV, the feasibility of this new approach is demonstrated. Therefore, whole genome RH maps will be soon available using this approach.

Whole genome RH maps

In Chapter IV, we showed that the new RH mapping method - genotyping by sequencing – has allowed the construction of RH linkage maps and the ordering and assigning of duck scaffolds along the chromosomes together with cytogenetic data. However, as mentioned in chapter IV, the CBS segmentation can assign breakpoints in a wrong sliding window or can fail to detect a breakage in some cases. This problem is difficult to solve due to the algorithm itself. To solve this problem we are developing a new segmentation program

mentioned in Chapter IV, which still needs to be improved due to its failure for the moment in detecting multiple breakages within one scaffold. For the moment, the CBS segmentation algorithm has a better performance and gives on average a higher number of markers on the robust maps. Therefore, until our new segmentation algorithm is improved, whole genome RH maps will be done using CBS segmentation. Most of map construction processes have their flaws and one must keep this in mind when using genome maps. It will always be necessary to compare maps obtained with different mapping methods: FISH, RH, genetic, BAC contigs… In addition, comparative mapping using a reference genome could require stronger evidence to suggest the marker order would be different from reference, therefore some false genotyping data could be compensated by the reference order. This is reflected by the detected rearrangements from the RH map of APL2, have been further confirmed by FISH experiments and previous cytogenetic data (Fillon et al. 2007; Skinner et al. 2009). In the short term, a genetic map containing 384 SNP markers is under construction in the laboratory by using the INRA GeneCan QTL resource mapping population. The resolution of this map will be lower than that of the RH map; but it will be built completely independently from external data, without using a reference genome. Any large-scale arrors in the RH map will be then detected.

Avian chromosome evolution

Avian chromosomes are well known to have highly conserved karyotypes and syntenies (Nanda et al. 2011; Shibusawa et al. 2004). Two-thirds of birds have a chromosome number in the order of 2n=74~86 (Griffin et al. 2007). The cross species painting experiment between chicken and nine other birds species belonging to 6 different orders diverged about 100MYA made by Guttenbach *et al* showed a striking conservation of synteny among those birds (Guttenbach et al. 2003). For the three birds having their genome sequence assembled into chromosomes: chicken, zebra finch and turkey, extensive studies showed that very few

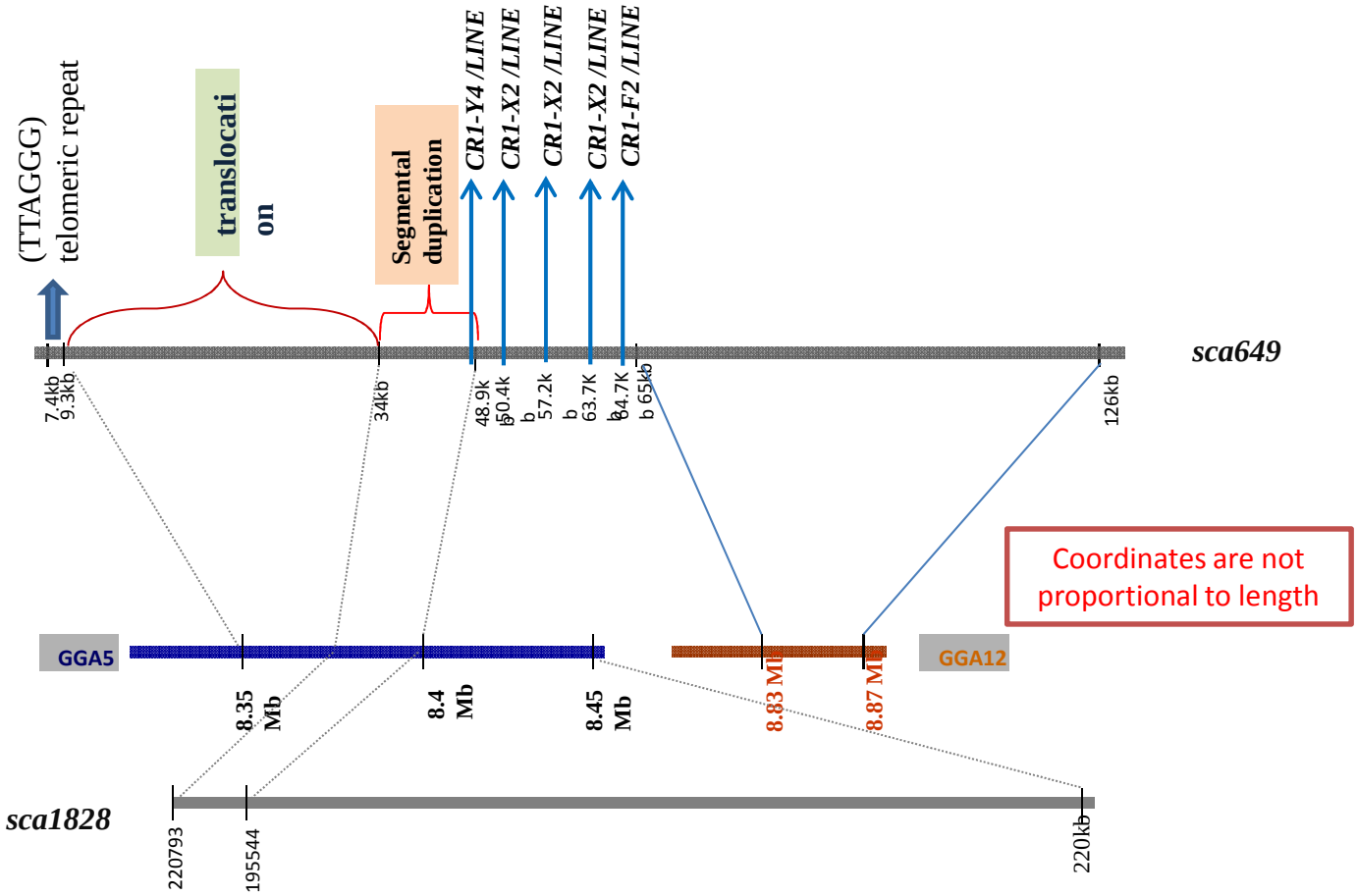


Figure V-1: A case of potential interchromosomal rearrangement between chicken and duck. RH linkage analysis showed that sca649 located on APL13, it is the first marker on the map (orientation: sca649_0 -> sca649_1). Sca1828 is the first marker on APL5 (orientation: sca1828_1 => sca1828_0) The repeats in the breakpoint region indicated by arrows were detected by RepeatMasker using chicken repeat data.

A scenario was proposed as follows: the first 34kb of sca649 perhaps was on a terminal on Chromome5, due to the segmental duplication, a non-allelic homologous recombination happened between Chromosome5 and Chromosome13, the 34kb fragment was exchanged to chromosme13.

interchromosomal rearrangements could be detected, most of which being fission or fusion events explaining the differences in chromosome number (Dalloul et al. 2010; Griffin et al. 2008; Reed et al. 2005; Reed et al. 2007; Stapley et al. 2008; Warren et al. 2010). Therefore for avian species, the intrachromosomal rearrangements may be the main driving force in speciation. There are only about 20~27 observed intrachromosomal rearrangement between turkey and chicken in which most are inversion despite 20 ~ 47 MYA divergence (Dimcheff et al. 2002; van Tuinen and Dyke 2004; Zhang et al. 2012), whereas there are 56 tentative inversions and 58 tentative translocation between chicken and zebra finch with an approximate 100MYA divergence (Pereira and Baker 2006; Volker et al. 2010; Warren et al. 2010). A recent study made by Skinner *et al* has compared the macrochromosomes of three sequenced birds, suggesting that about one-third of the chromosomal breakpoint regions may recur during avian evolution, from which the finding is also in agreement with their previous hypothesis that non-allelic homologous recombination (NAHR) hotspot drives genome evolution (Skinner and Griffin 2011).

The phylogenetic distance between duck and chicken is shorter than between chicken and zebra finch, with about 80 MYA for the former and 100MYA for the latter (Pereira and Baker 2006). Thus the duck genome will offer great insight and more evidence in bird chromosome evolution, with a number of expected chicken-duck rearrangements between the chicken- turkey and chicken-zebrafinch numbers. For example, a case we have found in Chapter III, sca649 would probably be an interchromosomal rearrangement which could be explained by this NAHR mechanism (demonstrated in Figure V-1). The RH maps of the three chromosomes, APL2, APL12 and APL22, suggest some intrachromosomal rearrangements among the four sequenced birds. These results can update the current comparative genomic data between duck and other birds; since to date, only rearrangements involving macrochromosomes have been identified between duck and chicken (Fillon et al. 2007; Skinner et al. 2009). The RH maps of APL2 and APL12 provide further evidence for

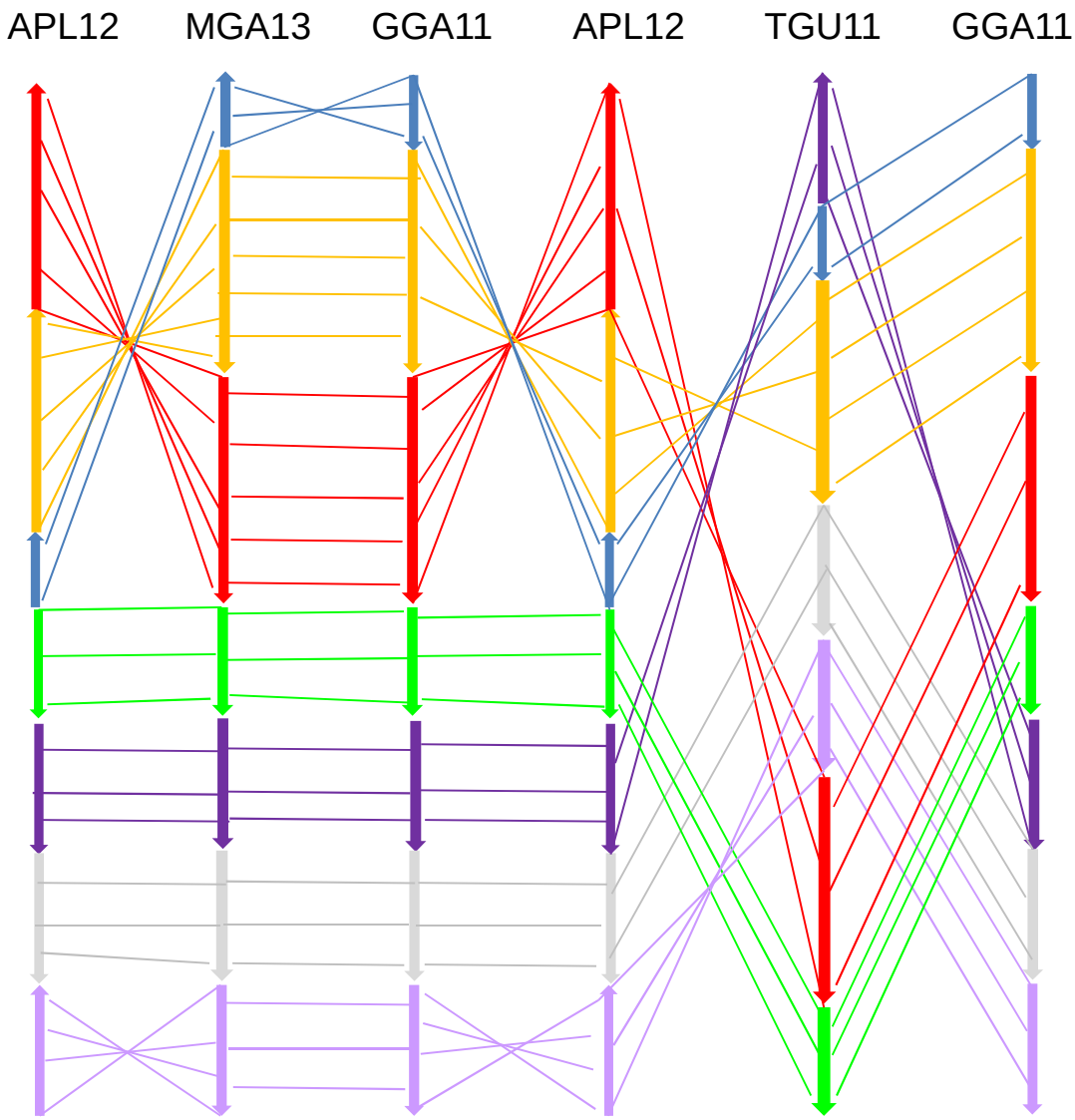


Figure V-2: rearrangements of APL12 compared with GGA11, TGU11 and MGA13. When comparing the 4 species, we could define 7 blocks of conserved synteny. When comparing APL12 with GGA11, there are two inversions, the largest of which was confirmed by FISH mapping. The rearrangements between APL12 and TGU11 are more complex. The number of rearrangements above are highly consistent with the divergence times, with the lower number between turkey and chicken, the highest between chicken/duck and zebra finch and an intermediate number between chicken and duck.

.

extensive intrachromosomal rearrangements in bird evolution (we exclude APL22 from our interpretations, due to the low number of markers on this map). We have defined 7 conserved synteny blocks (CSB) on APL12 **(Figure V-2)** and 18 CSB on APL2. These data suggests that chromosomal rearrangements detected are fewest between turkey and chicken, moderate between chicken and duck and highest between chicken/duck and zebra finch, in an agreement with phylogenetic data (van Tuinen and Hedges 2001). With the duck genome, some new evolutionary breakpoints which are not detected by comparison of chicken, turkey and zebra finch genome, could be therefore detected, e.g. two inversions on APL2 mentioned in Chapter IV. Also, the small inversion between GGA2 and TGU2 in the 2Mb region between positions 1Mb and 3.2Mb on GGA2 could be deduced as having happened in the Galliforme lineage (Skinner and Griffin 2011), as APL2 has the same order as zebra finch. It is interesting to note that most rearranged regions are close to the centromeres, suggesting that centromeres could play an important role in the rearrangements and perhaps in speciation. However, this observation will have to be confirmed by the comparative maps of the other chromosomes.

Observations on the composition of the genome around the avian breakpoints showed biases in repeat and GC % content (Gordon et al. 2007; Skinner and Griffin 2011). The construction of our whole genome RH maps will allow having more data on evolutionary breakpoint to confirm these observations and will participate in the reconstruction of the avian ancestral chromosomes.

The highly repeated gene: ATG4A

As we discussed in Chapter IV, we found the *ATG4A* gene to be highly redundant in the duck genome and the sequencing data shows that most copies do not seem truncated. This gene was not previously reported to be highly repeated in the three other sequenced birds. To check this, we performed FISH experiments using the *ATG4A* gene as a probe in chicken and

duck, confirming the gene might be only highly repeated in duck which is the only waterfowl among the sequenced birds. To further determine if this gene is highly repetitive in other waterfowl, we have checked preliminary sequencing data from Muscovy duck (A.Vignal, personnel communication). The average sequencing depth of the Muscovy duck genome in the data is about 30X whereas it is around 20,000 X for *ATG4A*, supposing this gene exists at a very high copy number in both species. As the divergence between common duck and Muscovy duck is about 20 MYA (http://www.timetree.org/), we speculate that the expansion of this gene must have happened before the divergence of the common duck and the Muscovy duck, but after that of chicken and duck about 80 MYA (van Tuinen and Dyke 2004).

The alignment of the reads from sequencing the duck radiation hybrids and the Muscovy duck suggest that this gene could still probably be active because indels and mismatches are not frequent in the reads. Interestingly, this gene in flanked by an LTR on one side only belonging to the GGLTR8B family, as determined by RepeatMasker (http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker). It is uncertain whether a second LTR exists and the gene was found on only one scaffold, meaning that the other copies were not assembled. However, it is possible that the gene could have expanded in the genome through an LTR machinery which is mainly a pathway through which genes are amplified in plants (Shirasu et al. 2000; Wicker et al. 2001). To understand whether this gene expansion is general in Anseriformes or restricted to ducks, two approaches could be taken. One is to perform FISH experiments on metaphases from different species. The other could be to perform a low coverage survey sequencing of these species, to check for an unusually high coverage of the gene in the sequencing data. To further investigate the possible role of this gene, additional data on transcript levels and functional data will be necessary..

Additional chromosomes in hybrids

The cytogenetic study in Chapter II showed that chicken and duck hybrids behaved similarly, with a variable number of chromosomes and a main integration pattern of donor cell DNA by forming additional chromosomes in hybrid cell. In some instances, the duck/chicken fragments could be inserted into the hamster genome or added on a chromosome end. The detailed process happening during the fusion is not known yet, but it is certain that donor chromosomes are broken in small fragments and randomly rescued to form additional chromosomes. It would be interesting to study whether there are some preferred motifs favored by the DNA repair machinery. It would be interesting to perform deep sequencing of a few hybrids using various inser-size libraries to investigate this.

When whole genome RH maps are available, additional exploration could be made to investigate whether some chromatin regions are more fragile to radiation. RH maps could allow us build virtual chromosomes, and then all the sequencing reads would be remapped to virtual chromosomes from which the breakpoint induced by radiation could be kept for further analysis.

Unraveling the smallest microchromosomes by Fluidigm Biomark qPCR

We have shown the power of using Fluidigm BioMark qPCR in RH mapping EST in Chapter III. The smallest microchromosomes in birds have some certain features causing difficulties in cloning and sequencing (ICGC. 2004; Morisson et al. 2007). Morisson *et al* has reported a strategy to construct RH linkage groups for the smallest microchromosomes in chicken (Morisson et al. 2007). Here we could suggest alternative complementary method to construct RH linkage groups by taking advantage of the Fluidigm BioMark qPCR.

Apply RH sequencing on other species

We have proved that 0.3X sequencing allows assignment of most duck scaffolds (considering the covered length) on RH linkage groups which are thereafter assigned to specific duck chromosomes. Since the number of species sequenced by NGS platforms keeps increasing, the corresponding genome assembly would be highly fragmented in the absence of an intermediate map, such as the panda genome (Li et al. 2010a). The importance of a genome assembly reaching chromosome level is important for comparative approach, reconstruction of ancestral genome, providing start point for assessment of gene expansion, contraction and adaptation, as illustrated by Lewin *et al* (Lewin et al. 2009). Thus, we suggest that the RH sequencing method could be applied to other species that are sequenced in by WGS. The most important feature is the mapping power (Θ), meaning the minimum distance (cR) for which breakages can be observed, which is determined by the number of hybrids and the retention of the panel (detailed in Chapter IV). Therefore, if the number of hybrids is constant, the mapping power could reach the highest value when retention is 50%, likewise, for a constant retention value, the more the hybrids, the higher the mapping power. Another parameter is the resolution which could be tailored by the radiation dose; the higher the dose, the high the resolution.

While applying this methodology on other species, several factors should be taken into account: (1) *de novo* genome assembly statistics; (2) average retention of the hybrid; (3) radiation dose; (4) hybrids number and (5) repeat content. If a genome assembly contains mostly contig/scaffold of small size; high radiation dose and higher sequencing depth would be required in order to observe breakages within smaller distances. In terms of retention, 50% would empower the mapping to the highest. However, it has proven almost impossible to obtain hybrids with such retention values. Moreover, retention values between 30% and 50% does not affect the resolution power, as reflected by the derivation of r*(1-r), which is less variable while close to r=0.5. Herein for the species having a genome assembly close to (or

slightly more fragmented than) duck, similar retention and radiation dose to what we have used is sufficient. As for the number of hybrids to sequence, two alternative approaches could be taken: either sequence less hybrids having higher retention or more hybrids with reasonably lower retention. This is also determined by the difficulties in obtaining hybrids and cost. Care should be taken for repeats which are very abundant in mammals and some fishes. The case study we made based on 0.3X sequencing benefits from the lower content in repetitive sequences that exists in birds. However and despite this, about one thirds of reads probably originating from duck DNA have been filtered to avoid false calling. Thus with a more repeat-rich species, higher sequencing depth and larger insert size for the paired reads would be desirable.

As sequencing cost drop down, we suggest that future project in RH sequencing could sequence at a higher depth, to avoid some false calling, as found in our data. We found some large scaffold with an unexpectedly low number of reads. These data correspond certainly to duck chromosome fragments present in a low percentage of cells in a hybrid. With deeper sequencing, such fragments may become easier to call, increasing the number of analyzable duck fragments and thus the retention values and mapping power. Contrariwise, some regions have unexpectedly high read numbers, possibly due to non-filtered repeats or to local locus amplification that could have happened following the stress in hybrids, such as shown in Chapter IV (Schimke 1984). Moreover, data from sequenced WGA hybrids could shed light on events happening when making other RH panel obtained by WGA such as sea bass (Guyon et al. 2010), rainbow trout (Y.Guiguen unpublished data), that the WGA panel needs a higher sequencing depth concerning the proportion of reads mapped on donor cell reduced. This might result from majority DNA contents being hamster origin (6Gb versus 200Mb from duck) that are more amplified in the hybrids.

Apart from sequencing WGA hybrids, we also sequenced some pooled hybrids which came from pooling two low retention hybrids. It is argued that pooling low retention hybrids

could function as having hybrids with a good retention (Lunetta et al. 1995). Thus we generated and sequenced 6 pooled hybrids, each composed of a mixture of two hybrids. Our data suggested that pooled hybrids have an average lower read coverage than a single non-pooled hybrid (data not shown). Thus, for species for which it is difficult to obtain hybrids, such as birds, fishes or perhaps reptilians, RH panel could be composed of pooled hybrids. However, in this case, sequencing the RH panel requires increasing the sequencing depth to compensate for the dilution of the donor cell DNA. Therefore, if pooling hybrids is a possible strategy for conventional genotyping by PCR, it has no great interest with the sequencing approach.

In conclusion, a higher sequencing depth is always desirable to avoid false calling or some biases resulting from heterogeneity of hybrid cell lines or for other reasons that we have discussed above. For the species that are planned to be sequenced, simulations could be done to optimize the sequencing coverage used for genome sequencing and RH sequencing to obtain a good assembly at a lower cost.

References

1991. *Microlivestock: Little-Known Small Animals with a Promising Economic Future*. The National Academies Press.

Adey, N.B., T.O. Tollefsbol, A.B. Sparks, M.H. Edgell, and C.A. Hutchison, 3rd. 1994. Molecular resurrection of an extinct ancestral promoter for mouse L1. *Proc Natl Acad Sci U S A* **91:** 1569-1573.

Adkins, R.M., E.L. Gelke, D. Rowe, and R.L. Honeycutt. 2001. Molecular phylogeny and divergence time estimates for major rodent groups: evidence from multiple genes. *Mol Biol Evol* **18:** 777-791.

Agate, R.J., M. Choe, and A.P. Arnold. 2004. Sex differences in structure and expression of the sex chromosome genes CHD1Z and CHD1W in zebra finches. *Mol Biol Evol* **21:** 384-396.

Agate, R.J., W. Grisham, J. Wade, S. Mann, J. Wingfield, C. Schanen, A. Palotie, and A.P. Arnold. 2003. Neural, not gonadal, origin of brain sex differences in a gynandromorphic finch. *Proc Natl Acad Sci U S A* **100:** 4873-4878.

Andreozzi, L., C. Federico, S. Motta, S. Saccone, A.L. Sazanova, A.A. Sazanov, A.F. Smirnov, S.A. Galkina, N.A. Lukina, A.V. Rodionov, N. Carels, and G. Bernardi. 2001. Compositional mapping of chicken chromosomes and identification of the gene-richest regions. *Chromosome Res* **9:** 521-532.

Ansorge, W.J. 2009. Next-generation DNA sequencing techniques. *N Biotechnol* **25:** 195-203.

Arnold, A.P., Y. Itoh, and E. Melamed. 2008. A bird's-eye view of sex chromosome dosage compensation. *Annu Rev Genomics Hum Genet* **9:** 109-127.

Augui, S., E.P. Nora, and E. Heard. 2011. Regulation of X-chromosome inactivation by the X-inactivation centre. *Nat Rev Genet* **12:** 429-442.

Avner, P., T. Bruls, I. Poras, L. Eley, S. Gas, P. Ruiz, M.V. Wiles, R. Sousa-Nunes, R. Kettleborough, A. Rana, J. Morissette, L. Bentley, M. Goldsworthy, A. Haynes, E. Herbert, L. Southam, H. Lehrach, J. Weissenbach, G. Manenti, P. Rodriguez-Tome, R. Beddington, S. Dunwoodie, and R.D. Cox. 2001. A radiation hybrid transcript map of the mouse genome. *Nat Genet* **29:** 194-200.

Barber, M.R., J.R. Aldridge, Jr., R.G. Webster, and K.E. Magor. 2010. Association of RIG-I with innate immunity of ducks to influenza. *Proc Natl Acad Sci U S A* **107:** 5913-5918.

Barnes, W.M. 1992. The fidelity of Taq polymerase catalyzing PCR is improved by an N-terminal deletion. *Gene* **112:** 29-35.

Bas, A., M. Swamy, L. Abeler-Dorner, G. Williams, D.J. Pang, S.D. Barbee, and A.C. Hayday. 2011. Butyrophilin-like 1 encodes an enterocyte protein that selectively regulates functional interactions with T lymphocytes. *Proc Natl Acad Sci U S A* **108:** 4376-4381.

Batellier, F., F. Marchal, M.F. Scheller, J. Gautron, N. Sellier, M. Taouis, C. Monbrun, A. Vignal, and J.P. Brillard. 2004. Sex ratios in mule duck embryos at various stages of incubation. *Theriogenology* **61:** 573-580.

Baverstock, P.R., M. Adams, R.W. Polkinghorne, and M. Gelder. 1982. A sex-linked enzyme in birds--Z-chromosome conservation but no dosage compensation. *Nature* **296:** 763-766.

Bed'Hom, B., P. Coullin, Z. Guillier-Gencik, S. Moulin, A. Bernheim, and V. Volobouev. 2003. Characterization of the atypical karyotype of the black-winged kite Elanus caeruleus (Falconiformes: Accipitridae) by means of classical and molecular cytogenetic techniques. *Chromosome Res* **11:** 335-343.

Benham, F., K. Hart, J. Crolla, M. Bobrow, M. Francavilla, and P.N. Goodfellow. 1989. A method for generating hybrids containing nonselected fragments of human chromosomes. *Genomics* **4:** 509-517.

Bentley, D.R. S. Balasubramanian H.P. Swerdlow G.P. Smith J. Milton C.G. Brown K.P. Hall D.J. Evers C.L. Barnes H.R. Bignell J.M. Boutell J. Bryant R.J. Carter R. Keira Cheetham A.J. Cox D.J. Ellis M.R. Flatbush N.A. Gormley S.J. Humphray L.J. Irving M.S. Karbelashvili S.M. Kirk H. Li X. Liu K.S. Maisinger L.J. Murray B. Obradovic T. Ost M.L. Parkinson M.R. Pratt I.M. Rasolonjatovo M.T. Reed R. Rigatti C. Rodighiero M.T. Ross A. Sabot S.V. Sankar A. Scally G.P. Schroth M.E. Smith V.P. Smith A. Spiridou P.E. Torrance S.S. Tzonev E.H. Vermaas K. Walter X. Wu L. Zhang M.D. Alam C. Anastasi I.C. Aniebo D.M. Bailey I.R. Bancarz S. Banerjee S.G. Barbour P.A. Baybayan V.A. Benoit K.F. Benson C. Bevis P.J. Black A. Boodhun J.S. Brennan J.A. Bridgham R.C. Brown A.A. Brown D.H. Buermann A.A. Bundu J.C. Burrows N.P. Carter N. Castillo E.C.M. Chiara S. Chang R. Neil Cooley N.R. Crake O.O. Dada K.D. Diakoumakos B. Dominguez-Fernandez D.J. Earnshaw U.C. Egbujor D.W. Elmore S.S. Etchin M.R. Ewan M. Fedurco L.J. Fraser K.V. Fuentes Fajardo W. Scott Furey D. George K.J. Gietzen C.P. Goddard G.S. Golda P.A. Granieri D.E. Green D.L. Gustafson N.F. Hansen K. Harnish C.D. Haudenschild N.I. Heyer M.M. Hims J.T. Ho A.M. Horgan K. Hoschler S. Hurwitz D.V. Ivanov M.Q. Johnson T. James T.A. Huw Jones G.D. Kang T.H. Kerelska A.D. Kersey I. Khrebtukova A.P. Kindwall Z. Kingsbury P.I. Kokko-Gonzales A. Kumar M.A. Laurent C.T. Lawley S.E. Lee X. Lee A.K. Liao J.A. Loch M. Lok S. Luo R.M. Mammen J.W. Martin P.G. McCauley P. McNitt P. Mehta K.W. Moon J.W. Mullens T. Newington Z. Ning B. Ling Ng S.M. Novo M.J. O'Neill M.A. Osborne A. Osnowski O. Ostadan L.L. Paraschos L. Pickering A.C. Pike D. Chris Pinkard D.P. Pliskin J. Podhasky V.J. Quijano C. Raczy V.H. Rae S.R. Rawlings A. Chiva Rodriguez P.M. Roe J. Rogers M.C. Rogert Bacigalupo N. Romanov A. Romieu R.K. Roth N.J. Rourke S.T. Ruediger E. Rusman R.M. Sanches-Kuiper M.R. Schenker J.M. Seoane R.J. Shaw M.K. Shiver S.W. Short N.L. Sizto J.P. Sluis M.A. Smith J. Ernest Sohna Sohna E.J.

Spence K. Stevens N. Sutton L. Szajkowski C.L. Tregidgo G. Turcatti S. Vandevondele Y. Verhovsky S.M. Virk S. Wakelin G.C. Walcott J. Wang G.J. Worsley J. Yan L. Yau M. Zuerlein J.C. Mullikin M.E. Hurles N.J. McCooke J.S. West F.L. Oaks P.L. Lundberg D. Klenerman R. Durbin and A.J. Smith. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456:** 53-59.

Berthier-Schaad, Y., W.H. Kao, J. Coresh, L. Zhang, R.G. Ingersoll, R. Stephens, and M.W. Smith. 2007. Reliability of high-throughput genotyping of whole genome amplified DNA in SNP genotyping studies. *Electrophoresis* **28:** 2812-2817.

Bitgood, J.J. and R.G.J. Somes. 1993. Gene map of the chicken (Gallus gallus). In *Genetic Maps' 6th edition* (ed. S. O'Brien), pp. 4.333-334.342, Cold Spring Harbor Laboratory Press.

Braslavsky, I., B. Hebert, E. Kartalov, and S.R. Quake. 2003. Sequence information can be obtained from single DNA molecules. *Proc Natl Acad Sci U S A* **100:** 3960-3964.

Brenner, S., M. Johnson, J. Bridgham, G. Golda, D.H. Lloyd, D. Johnson, S. Luo, S. McCurdy, M. Foy, M. Ewan, R. Roth, D. George, S. Eletr, G. Albrecht, E. Vermaas, S.R. Williams, K. Moon, T. Burcham, M. Pallas, R.B. DuBridge, J. Kirchner, K. Fearon, J. Mao, and K. Corcoran. 2000. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat Biotechnol* **18:** 630-634.

Brown, C.J., A. Ballabio, J.L. Rupert, R.G. Lafreniere, M. Grompe, R. Tonlorenzi, and H.F. Willard. 1991. A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* **349:** 38-44.

Burt, D.W. 2002. Origin and evolution of avian microchromosomes. *Cytogenet Genome Res* **96:** 97-112.

Burt, D.W., C. Bruley, I.C. Dunn, C.T. Jones, A. Ramage, A.S. Law, D.R. Morrice, I.R. Paton, J. Smith, D. Windsor, A. Sazanov, R. Fries, and D. Waddington. 1999. The dynamics of chromosome evolution in birds and mammals. *Nature* **402:** 411-413.

Cassingena, R., C. Chany, M. Vignal, H. Suarez, S. Estrade, and P. Lazar. 1971. Use of monkey-mouse hybrid cells for the study of the cellular regualtion of interferon production and action. *Proc Natl Acad Sci U S A* **68:** 580-584.

Charlesworth, D. and B. Charlesworth. 2005. Sex chromosomes: evolution of the weird and wonderful. *Curr Biol* **15:** R129-131.

Chen, F.C. and W.H. Li. 2001. Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. *Am J Hum Genet* **68:** 444-456.

Chen, X., R.J. Agate, Y. Itoh, and A.P. Arnold. 2005. Sexually dimorphic expression of trkB, a Z-linked gene, in early posthatch zebra finch brain. *Proc Natl Acad Sci U S A* **102:** 7730-7735.

Cherry, P. and T. Morris. 2008. *Domestic Duck Production Science and Practice*. CAB International.

Chojnowski, J.L., R.T. Kimball, and E.L. Braun. 2008. Introns outperform exons in analyses of basal avian phylogeny using clathrin heavy chain genes. *Gene* **410:** 89-96.

Chowdhary, B.P. and T. Raudsepp. 2000. HSA4 and GGA4: remarkable conservation despite 300-Myr divergence. *Genomics* **64:** 102-105.

Chowdhary, B.P., T. Raudsepp, D. Honeycutt, E.K. Owens, F. Piumi, G. Guerin, T.C. Matise, S.R. Kata, J.E. Womack, and L.C. Skow. 2002. Construction of a 5000(rad) whole-genome radiation hybrid panel in the horse and generation of a comprehensive and comparative map for ECA11. *Mamm Genome* **13:** 89-94.

Chubb, A.L. 2004. New nuclear evidence for the oldest divergence among neognath birds: the phylogenetic utility of ZENK (i). *Mol Phylogenet Evol* **30:** 140-151.

Clark, T.A., K.E. Spittle, S.W. Turner, and J. Korlach. 2011. Direct detection and sequencing of damaged DNA bases. *Genome Integr* **2:** 10.

Clayton, D.F., C.N. Balakrishnan, and S.E. London. 2009. Integrating genomes, brain and behavior in the study of songbirds. *Curr Biol* **19:** R865-873.

Clayton, G.A. 1984. Common duck. In *Evolution of domesticated animals* (ed. I.L. Mason), pp. 334-339, London.

Clemson, C.M., J.A. McNeil, H.F. Willard, and J.B. Lawrence. 1996. XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. *J Cell Biol* **132:** 259-275.

Conrad, T. and A. Akhtar. 2012. Dosage compensation in Drosophila melanogaster: epigenetic fine-tuning of chromosome-wide transcription. *Nat Rev Genet* **13:** 123-134.

Consortium, I.C.G. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432:** 695-716.

Cox, D.R., M. Burmeister, E.R. Price, S. Kim, and R.M. Myers. 1990. Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. *Science* **250:** 245-250.

Dalloul, R.A., J.A. Long, A.V. Zimin, L. Aslam, K. Beal, A. Blomberg Le, P. Bouffard, D.W. Burt, O. Crasta, R.P. Crooijmans, K. Cooper, R.A. Coulombe, S. De, M.E. Delany, J.B. Dodgson, J.J. Dong, C. Evans, K.M. Frederickson, P. Flicek, L. Florea, O. Folkerts, M.A. Groenen, T.T. Harkins, J. Herrero, S. Hoffmann, H.J. Megens, A. Jiang, P. de Jong, P. Kaiser, H. Kim, K.W. Kim, S. Kim, D. Langenberger, M.K. Lee, T. Lee, S. Mane, G. Marcais, M. Marz, A.P. McElroy, T. Modise, M. Nefedov, C. Notredame, I.R. Paton, W.S. Payne, G. Pertea, D. Prickett, D. Puiu, D. Qioa, E. Raineri, M. Ruffier, S.L. Salzberg, M.C. Schatz, C. Scheuring, C.J. Schmidt, S. Schroeder, S.M. Searle, E.J. Smith, J. Smith, T.S. Sonstegard, P.F. Stadler, H. Tafer, Z.J. Tu, C.P. Van Tassell, A.J. Vilella, K.P. Williams, J.A. Yorke, L. Zhang, H.B. Zhang, X. Zhang, Y.

Zhang, and K.M. Reed. 2010. Multi-platform next-generation sequencing of the domestic turkey (Meleagris gallopavo): genome assembly and analysis. *PLoS Biol* **8**.

Dean, F.B., S. Hosono, L. Fang, X. Wu, A.F. Faruqi, P. Bray-Ward, Z. Sun, Q. Zong, Y. Du, J. Du, M. Driscoll, W. Song, S.F. Kingsmore, M. Egholm, and R.S. Lasken. 2002. Comprehensive human genome amplification using multiple displacement amplification. *Proc Natl Acad Sci U S A* **99:** 5261-5266.

Dear, P. 2001. Genome mapping. In *ENCYCLOPEDIA OF LIFE SCIENCES*.

Dehal, P., P. Predki, A.S. Olsen, A. Kobayashi, P. Folta, S. Lucas, M. Land, A. Terry, C.L. Ecale Zhou, S. Rash, Q. Zhang, L. Gordon, J. Kim, C. Elkin, M.J. Pollard, P. Richardson, D. Rokhsar, E. Uberbacher, T. Hawkins, E. Branscomb, and L. Stubbs. 2001. Human chromosome 19 and related regions in mouse: conservative and lineage-specific evolution. *Science* **293:** 104-111.

Denjean, B., A. Ducos, A. Darre, A. Pinton, A. Seguela, H. Berland, and Blanc. 1997. Caryotypes des canards commun (Anas platyrhynchos), Barbarie (Cairina moschata) et de leur hybride. *Revue. Méd. Vet***:** 695.

Derjusheva, S., A. Kurganova, F. Habermann, and E. Gaginskaya. 2004. High chromosome conservation detected by comparative chromosome painting in chicken, pigeon and passerine birds. *Chromosome Res* **12:** 715-723.

Dimcheff, D.E., S.V. Drovetski, and D.P. Mindell. 2002. Phylogeny of Tetraoninae and other galliform birds using mitochondrial 12S and ND2 genes. *Mol Phylogenet Evol* **24:** 203-215.

Dohm, J.C., C. Lottaz, T. Borodina, and H. Himmelbauer. 2008. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res* **36:** e105.

Douaud, M., K. Feve, M. Gerus, V. Fillon, S. Bardes, D. Gourichon, D.A. Dawson, O. Hanotte, T. Burke, F. Vignoles, M. Morisson, M. Tixier-Boichard, A. Vignal, and F. Pitel. 2008. Addition of the microchromosome GGA25 to the chicken genome sequence assembly through radiation hybrid and genetic mapping. *BMC Genomics* **9:** 129.

Doupe, A.J. and P.K. Kuhl. 1999. Birdsong and human speech: common themes and mechanisms. *Annu Rev Neurosci* **22:** 567-631.

Duan, L., L. Campitelli, X.H. Fan, Y.H. Leung, D. Vijaykrishna, J.X. Zhang, I. Donatelli, M. Delogu, K.S. Li, E. Foni, C. Chiapponi, W.L. Wu, H. Kai, R.G. Webster, K.F. Shortridge, J.S. Peiris, G.J. Smith, H. Chen, and Y. Guan. 2007. Characterization of low-pathogenic H5 subtype influenza viruses from Eurasia: implications for the origin of highly pathogenic H5N1 viruses. *J Virol* **81:** 7529-7539.

Echard G., G.J., Gillois M. 1984. Localisation des gène MPI , PKM2,NP sur le chromosome3 de porc (Sus scrofa L.) et analyse cytogénétique d'une lignée de hamster chinois issuede la DON (Wg3h). *Genet. Sel. Evol* **16:** 261-270.

Eichler, E.E. and D. Sankoff. 2003. Structural dynamics of eukaryotic chromosome evolution. *Science* **301:** 793-797.

Eid, J., A. Fehr, J. Gray, K. Luong, J. Lyle, G. Otto, P. Peluso, D. Rank, P. Baybayan, B. Bettman, A. Bibillo, K. Bjornson, B. Chaudhuri, F. Christians, R. Cicero, S. Clark, R. Dalal, A. Dewinter, J. Dixon, M. Foquet, A. Gaertner, P. Hardenbol, C. Heiner, K. Hester, D. Holden, G. Kearns, X. Kong, R. Kuse, Y. Lacroix, S. Lin, P. Lundquist, C. Ma, P. Marks, M. Maxham, D. Murphy, I. Park, T. Pham, M. Phillips, J. Roy, R. Sebra, G. Shen, J. Sorenson, A. Tomaney, K. Travers, M. Trulson, J. Vieceli, J. Wegener, D. Wu, A. Yang, D. Zaccarin, P. Zhao, F. Zhong, J. Korlach, and S. Turner. 2009. Real-time DNA sequencing from single polymerase molecules. *Science* **323:** 133-138.

Ericson, P.G., C.L. Anderson, T. Britton, A. Elzanowski, U.S. Johansson, M. Kallersjo, J.I. Ohlson, T.J. Parsons, D. Zuccon, and G. Mayr. 2006. Diversification of Neoaves: integration of molecular sequence data and fossils. *Biol Lett* **2:** 543-547.

Faraut, T., S. de Givry, C. Hitte, Y. Lahbib-Mansais, M. Morisson, D. Milan, T. Schiex, B. Servin, A. Vignal, F. Galibert, and M. Yerle. 2009. Contribution of radiation hybrids to genome mapping in domestic animals. *Cytogenet Genome Res* **126:** 21-33.

Fedurco, M., A. Romieu, S. Williams, I. Lawrence, and G. Turcatti. 2006. BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. *Nucleic Acids Res* **34:** e22.

Fillon, V., M. Vignoles, R.P. Crooijmans, M.A. Groenen, R. Zoorob, and A. Vignal. 2007. FISH mapping of 57 BAC clones reveals strong conservation of synteny between Galliformes and Anseriformes. *Anim Genet* **38:** 303-307.

Flusberg, B.A., D.R. Webster, J.H. Lee, K.J. Travers, E.C. Olivares, T.A. Clark, J. Korlach, and S.W. Turner. 2010. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods* **7:** 461-465.

Foquet, M., K.T. Samiee, X. Kong, B. Chaudhuri, P. Lundquist, S.W. Turner, J. Freudenthal, and D.B. Roitman. 2008. Improved fabrication of zero-mode waveguides for single-molecule detection. *J. Appl. Phys* **103**.

Fridolfsson, A.K., H. Cheng, N.G. Copeland, N.A. Jenkins, H.C. Liu, T. Raudsepp, T. Woodage, B. Chowdhary, J. Halverson, and H. Ellegren. 1998. Evolution of the avian sex chromosomes from an ancestral pair of autosomes. *Proc Natl Acad Sci U S A* **95:** 8147-8152.

Geisler, R., G.J. Rauch, H. Baier, F. van Bebber, L. Bross, M.P. Dekens, K. Finger, C. Fricke, M.A. Gates, H. Geiger, S. Geiger-Rudolph, D. Gilmour, S. Glaser, L. Gnugge, H. Habeck, K. Hingst, S. Holley, J. Keenan, A. Kirn, H. Knaut, D. Lashkari, F. Maderspacher, U. Martyn, S. Neuhauss, C. Neumann, T. Nicolson, F. Pelegri, R. Ray, J.M. Rick, H. Roehl, T. Roeser, H.E. Schauerte, A.F. Schier, U. Schonberger, H.B. Schonthaler, S. Schulte-Merker, C. Seydler, W.S. Talbot, C. Weiler, C. Nusslein-

Volhard, and P. Haffter. 1999. A radiation hybrid map of the zebrafish genome. *Nat Genet* **23:** 86-89.

Gibbs, R.A. G.M. Weinstock M.L. Metzker D.M. Muzny E.J. Sodergren S. Scherer G. Scott D. Steffen K.C. Worley P.E. Burch G. Okwuonu S. Hines L. Lewis C. DeRamo O. Delgado S. Dugan-Rocha G. Miner M. Morgan A. Hawes R. Gill Celera R.A. Holt M.D. Adams P.G. Amanatides H. Baden-Tillson M. Barnstead S. Chin C.A. Evans S. Ferriera C. Fosler A. Glodek Z. Gu D. Jennings C.L. Kraft T. Nguyen C.M. Pfannkoch C. Sitter G.G. Sutton J.C. Venter T. Woodage D. Smith H.M. Lee E. Gustafson P. Cahill A. Kana L. Doucette-Stamm K. Weinstock K. Fechtel R.B. Weiss D.M. Dunn E.D. Green R.W. Blakesley G.G. Bouffard P.J. De Jong K. Osoegawa B. Zhu M. Marra J. Schein I. Bosdet C. Fjell S. Jones M. Krzywinski C. Mathewson A. Siddiqui N. Wye J. McPherson S. Zhao C.M. Fraser J. Shetty S. Shatsman K. Geer Y. Chen S. Abramzon W.C. Nierman P.H. Havlak R. Chen K.J. Durbin A. Egan Y. Ren X.Z. Song B. Li Y. Liu X. Qin S. Cawley A.J. Cooney L.M. D'Souza K. Martin J.Q. Wu M.L. Gonzalez-Garay A.R. Jackson K.J. Kalafus M.P. McLeod A. Milosavljevic D. Virk A. Volkov D.A. Wheeler Z. Zhang J.A. Bailey E.E. Eichler E. Tuzun E. Birney E. Mongin A. Ureta-Vidal C. Woodwark E. Zdobnov P. Bork M. Suyama D. Torrents M. Alexandersson B.J. Trask J.M. Young H. Huang H. Wang H. Xing S. Daniels D. Gietzen J. Schmidt K. Stevens U. Vitt J. Wingrove F. Camara M. Mar Alba J.F. Abril R. Guigo A. Smit I. Dubchak E.M. Rubin O. Couronne A. Poliakov N. Hubner D. Ganten C. Goesele O. Hummel T. Kreitler Y.A. Lee J. Monti H. Schulz H. Zimdahl H. Himmelbauer H. Lehrach H.J. Jacob S. Bromberg J. Gullings-Handley M.I. Jensen-Seaman A.E. Kwitek J. Lazar D. Pasko P.J. Tonellato S. Twigger C.P. Ponting J.M. Duarte S. Rice L. Goodstadt S.A. Beatson R.D. Emes E.E. Winter C. Webber P. Brandt G. Nyakatura M. Adetobi F. Chiaromonte L. Elnitski P. Eswara R.C. Hardison M. Hou D. Kolbe K. Makova W. Miller A. Nekrutenko C. Riemer S. Schwartz J. Taylor S. Yang Y. Zhang K. Lindpaintner T.D. Andrews M. Caccamo M. Clamp L. Clarke V. Curwen R. Durbin E. Eyras S.M. Searle G.M. Cooper S. Batzoglou M. Brudno A. Sidow E.A. Stone B.A. Payseur G. Bourque C. Lopez-Otin X.S. Puente K. Chakrabarti S. Chatterji C. Dewey L. Pachter N. Bray V.B. Yap A. Caspi G. Tesler P.A. Pevzner D. Haussler K.M. Roskin R. Baertsch H. Clawson T.S. Furey A.S. Hinrichs D. Karolchik W.J. Kent K.R. Rosenbloom H. Trumbower M. Weirauch D.N. Cooper P.D. Stenson B. Ma M. Brent M. Arumugam D. Shteynberg R.R. Copley M.S. Taylor H. Riethman U. Mudunuri J. Peterson M. Guyer A. Felsenfeld S. Old S. Mockrin and F. Collins. 2004. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428:** 493-521.

Goodfellow, P.J., S. Povey, H.A. Nevanlinna, and P.N. Goodfellow. 1990. Generation of a panel of somatic cell hybrids containing unselected fragments of human chromosome

10 by X-ray irradiation and cell fusion: application to isolating the MEN2A region in hybrid cells. *Somat Cell Mol Genet* **16:** 163-171.

Gordon, L., S. Yang, M. Tran-Gyamfi, D. Baggott, M. Christensen, A. Hamilton, R. Crooijmans, M. Groenen, S. Lucas, I. Ovcharenko, and L. Stubbs. 2007. Comparative analysis of chicken chromosome 28 provides new clues to the evolutionary fragility of gene-rich vertebrate regions. *Genome Res* **17:** 1603-1613.

Gorman, M. and B.S. Baker. 1994. How flies make one equal two: dosage compensation in Drosophila. *Trends Genet* **10:** 376-380.

Goss, S.J. and H. Harris. 1975. New method for mapping genes in human chromosomes. *Nature* **255:** 680-684.

Green, E.D. 2001. Strategies for the systematic sequencing of complex genomes. *Nat Rev Genet* **2:** 573-583.

Gregory, S., C. Soderlund, and A. Coulson. 1996. Contig assembly by fingerprinting. In *Genome Mapping: A Practical Approach.* (ed. P. Dear). Oxford University Press.

Gregory, S.G., M. Sekhon, J. Schein, S. Zhao, K. Osoegawa, C.E. Scott, R.S. Evans, P.W. Burridge, T.V. Cox, C.A. Fox, R.D. Hutton, I.R. Mullenger, K.J. Phillips, J. Smith, J. Stalker, G.J. Threadgold, E. Birney, K. Wylie, A. Chinwalla, J. Wallis, L. Hillier, J. Carter, T. Gaige, S. Jaeger, C. Kremitzki, D. Layman, J. Maas, R. McGrane, K. Mead, R. Walker, S. Jones, M. Smith, J. Asano, I. Bosdet, S. Chan, S. Chittaranjan, R. Chiu, C. Fjell, D. Fuhrmann, N. Girn, C. Gray, R. Guin, L. Hsiao, M. Krzywinski, R. Kutsche, S.S. Lee, C. Mathewson, C. McLeavy, S. Messervier, S. Ness, P. Pandoh, A.L. Prabhu, P. Saeedi, D. Smailus, L. Spence, J. Stott, S. Taylor, W. Terpstra, M. Tsai, J. Vardy, N. Wye, G. Yang, S. Shatsman, B. Ayodeji, K. Geer, G. Tsegaye, A. Shvartsbeyn, E. Gebregeorgis, M. Krol, D. Russell, L. Overton, J.A. Malek, M. Holmes, M. Heaney, J. Shetty, T. Feldblyum, W.C. Nierman, J.J. Catanese, T. Hubbard, R.H. Waterston, J. Rogers, P.J. de Jong, C.M. Fraser, M. Marra, J.D. McPherson, and D.R. Bentley. 2002. A physical map of the mouse genome. *Nature* **418:** 743-750.

Griffin, D.K., L.B. Robertson, H.G. Tempest, and B.M. Skinner. 2007. The evolution of the avian genome as revealed by comparative molecular cytogenetics. *Cytogenet Genome Res* **117:** 64-77.

Griffin, D.K., L.B. Robertson, H.G. Tempest, A. Vignal, V. Fillon, R.P. Crooijmans, M.A. Groenen, S. Deryusheva, E. Gaginskaya, W. Carre, D. Waddington, R. Talbot, M. Volker, J.S. Masabanda, and D.W. Burt. 2008. Whole genome comparative studies between chicken and turkey and their implications for avian genome evolution. *BMC Genomics* **9:** 168.

Groth, J.G. and G.F. Barrowclough. 1999. Basal divergences in birds and the phylogenetic utility of the nuclear RAG-1 gene. *Mol Phylogenet Evol* **12:** 115-123.

Grutzner, F., E. Zend-Ajusch, K. Stout, S. Munsche, A. Niveleau, I. Nanda, M. Schmid, and T. Haaf. 2001. Chicken microchromosomes are hypermethylated and can be identified by specific painting probes. *Cytogenet Cell Genet* **93:** 265-269.

Grzeschik, K.H., P.W. Allderdice, A. Grzeschik, J.M. Opitz, O.J. Miller, and M. Siniscalco. 1972. Cytological mapping of human X-linked genes by use of somatic cell hybrids involving an X-autosome translocation (mouse-hamster-human X-linked markers). *Proc Natl Acad Sci U S A* **69:** 69-73.

Guttenbach, M., I. Nanda, W. Feichtinger, J.S. Masabanda, D.K. Griffin, and M. Schmid. 2003. Comparative chromosome painting of chicken autosomal paints 1-9 in nine different bird species. *Cytogenet Genome Res* **103:** 173-184.

Guyon, R., F. Senger, M. Rakotomanga, N. Sadequi, F.A. Volckaert, C. Hitte, and F. Galibert. A radiation hybrid map of the European sea bass (Dicentrarchus labrax) based on 1581 markers: Synteny analysis with model fish genomes. *Genomics* **96:** 228-238.

Guyon, R., F. Senger, M. Rakotomanga, N. Sadequi, F.A. Volckaert, C. Hitte, and F. Galibert. 2010. A radiation hybrid map of the European sea bass (Dicentrarchus labrax) based on 1581 markers: Synteny analysis with model fish genomes. *Genomics* **96:** 228-238.

Guyon, R., F. Senger, M. Rakotomanga, N. Sadequi, F.A. Volckaert, C. Hitte, and F. Galibert. 2011. A radiation hybrid map of the European sea bass (Dicentrarchus labrax) based on 1581 markers: Synteny analysis with model fish genomes. *Genomics* **96:** 228-238.

Gyapay, G., K. Schmitt, C. Fizames, H. Jones, N. Vega-Czarny, D. Spillett, D. Muselet, J.F. Prud'homme, C. Dib, C. Auffray, J. Morissette, J. Weissenbach, and P.N. Goodfellow. 1996. A radiation hybrid map of the human genome. *Hum Mol Genet* **5:** 339-346.

Haas, N.B., J.M. Grabowski, J. North, J.V. Moran, H.H. Kazazian, and J.B. Burch. 2001. Subfamilies of CR1 non-LTR retrotransposons have different 5'UTR sequences but are otherwise conserved. *Gene* **265:** 175-183.

Habermann, F.A., M. Cremer, J. Walter, G. Kreth, J. von Hase, K. Bauer, J. Wienberg, C. Cremer, T. Cremer, and I. Solovei. 2001. Arrangements of macro- and microchromosomes in chicken cells. *Chromosome Res* **9:** 569-584.

Hackett, S.J., R.T. Kimball, S. Reddy, R.C. Bowie, E.L. Braun, M.J. Braun, J.L. Chojnowski, W.A. Cox, K.L. Han, J. Harshman, C.J. Huddleston, B.D. Marks, K.J. Miglia, W.S. Moore, F.H. Sheldon, D.W. Steadman, C.C. Witt, and T. Yuri. 2008. A phylogenomic study of birds reveals their evolutionary history. *Science* **320:** 1763-1768.

Hamasima, N., H. Suzuki, A. Mikawa, T. Morozumi, G. Plastow, and T. Mitsuhashi. 2003. Construction of a new porcine whole-genome framework map using a radiation hybrid panel. *Anim Genet* **34:** 216-220.

Harismendy, O., P.C. Ng, R.L. Strausberg, X. Wang, T.B. Stockwell, K.Y. Beeson, N.J. Schork, S.S. Murray, E.J. Topol, S. Levy, and K.A. Frazer. 2009. Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biol* **10:** R32.

Harris, T.D., P.R. Buzby, H. Babcock, E. Beer, J. Bowers, I. Braslavsky, M. Causey, J. Colonell, J. Dimeo, J.W. Efcavitch, E. Giladi, J. Gill, J. Healy, M. Jarosz, D. Lapen, K. Moulton, S.R. Quake, K. Steinmann, E. Thayer, A. Tyurina, R. Ward, H. Weiss, and Z. Xie. 2008. Single-molecule DNA sequencing of a viral genome. *Science* **320:** 106-109.

Heard, E. and C.M. Disteche. 2006. Dosage compensation in mammals: fine-tuning the expression of the X chromosome. *Genes Dev* **20:** 1848-1867.

Hitte, C., J. Madeoy, E.F. Kirkness, C. Priat, T.D. Lorentzen, F. Senger, D. Thomas, T. Derrien, C. Ramirez, C. Scott, G. Evanno, B. Pullar, E. Cadieu, V. Oza, K. Lourgant, D.B. Jaffe, S. Tacher, S. Dreano, N. Berkova, C. Andre, P. Deloukas, C. Fraser, K. Lindblad-Toh, E.A. Ostrander, and F. Galibert. 2005. Facilitating genome navigation: survey sequencing and dense radiation-hybrid gene mapping. *Nat Rev Genet* **6:** 643-648.

Hou, Y., L. Song, P. Zhu, B. Zhang, Y. Tao, X. Xu, F. Li, K. Wu, J. Liang, D. Shao, H. Wu, X. Ye, C. Ye, R. Wu, M. Jian, Y. Chen, W. Xie, R. Zhang, L. Chen, X. Liu, X. Yao, H. Zheng, C. Yu, Q. Li, Z. Gong, M. Mao, X. Yang, L. Yang, J. Li, W. Wang, Z. Lu, N. Gu, G. Laurie, L. Bolund, K. Kristiansen, J. Wang, H. Yang, Y. Li, and X. Zhang. 2012. Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm. *Cell* **148:** 873-885.

Huang, S., R. Li, Z. Zhang, L. Li, X. Gu, W. Fan, W.J. Lucas, X. Wang, B. Xie, P. Ni, Y. Ren, H. Zhu, J. Li, K. Lin, W. Jin, Z. Fei, G. Li, J. Staub, A. Kilian, E.A. van der Vossen, Y. Wu, J. Guo, J. He, Z. Jia, G. Tian, Y. Lu, J. Ruan, W. Qian, M. Wang, Q. Huang, B. Li, Z. Xuan, J. Cao, Asan, Z. Wu, J. Zhang, Q. Cai, Y. Bai, B. Zhao, Y. Han, Y. Li, X. Li, S. Wang, Q. Shi, S. Liu, W.K. Cho, J.Y. Kim, Y. Xu, K. Heller-Uszynska, H. Miao, Z. Cheng, S. Zhang, J. Wu, Y. Yang, H. Kang, M. Li, H. Liang, X. Ren, Z. Shi, M. Wen, M. Jian, H. Yang, G. Zhang, Z. Yang, R. Chen, L. Ma, H. Liu, Y. Zhou, J. Zhao, X. Fang, L. Fang, D. Liu, H. Zheng, Y. Zhang, N. Qin, Z. Li, G. Yang, S. Yang, L. Bolund, K. Kristiansen, S. Li, X. Zhang, J. Wang, R. Sun, B. Zhang, S. Jiang, and Y. Du. 2009. The genome of the cucumber, Cucumis sativus L. *Nat Genet* **41:** 1275-1281.

Huang, Y., C.S. Haley, S. Hu, J. Hao, C. Wu, and N. Li. 2007a. Detection of quantitative trait loci for body weights and conformation traits in Beijing ducks. *Anim Genet* **38:** 525-526.

Huang, Y., C.S. Haley, F. Wu, S. Hu, J. Hao, C. Wu, and N. Li. 2007b. Genetic mapping of quantitative trait loci affecting carcass and meat quality traits in Beijing ducks (Anas platyrhynchos). *Anim Genet* **38:** 114-119.

Huang, Y., Y. Zhao, C.S. Haley, S. Hu, J. Hao, C. Wu, and N. Li. 2006. A genetic and cytogenetic map for the duck (Anas platyrhynchos). *Genetics* **173:** 287-296.

Hughes, A.L. and M.K. Hughes. 1995. Small genomes for better flyers. *Nature* **377:** 391.

Hukriede, N., D. Fisher, J. Epstein, L. Joly, P. Tellis, Y. Zhou, B. Barbazuk, K. Cox, L. Fenton-Noriega, C. Hersey, J. Miles, X. Sheng, A. Song, R. Waterman, S.L. Johnson,

I.B. Dawid, M. Chevrette, L.I. Zon, J. McPherson, and M. Ekker. 2001. The LN54 radiation hybrid map of zebrafish expressed sequences. *Genome Res* **11:** 2127-2132.

Hukriede, N.A., L. Joly, M. Tsang, J. Miles, P. Tellis, J.A. Epstein, W.B. Barbazuk, F.N. Li, B. Paw, J.H. Postlethwait, T.J. Hudson, L.I. Zon, J.D. McPherson, M. Chevrette, I.B. Dawid, S.L. Johnson, and M. Ekker. 1999. Radiation hybrid mapping of the zebrafish genome. *Proc Natl Acad Sci U S A* **96:** 9745-9750.

Hulse-Post, D.J., J. Franks, K. Boyd, R. Salomon, E. Hoffmann, H.L. Yen, R.J. Webby, D. Walker, T.D. Nguyen, and R.G. Webster. 2007. Molecular changes in the polymerase genes (PA and PB1) associated with high pathogenicity of H5N1 influenza virus in mallard ducks. *J Virol* **81:** 8515-8524.

Hulse-Post, D.J., K.M. Sturm-Ramirez, J. Humberd, P. Seiler, E.A. Govorkova, S. Krauss, C. Scholtissek, P. Puthavathana, C. Buranathai, T.D. Nguyen, H.T. Long, T.S. Naipospos, H. Chen, T.M. Ellis, Y. Guan, J.S. Peiris, and R.G. Webster. 2005. Role of domestic ducks in the propagation and biological evolution of highly pathogenic H5N1 influenza viruses in Asia. *Proc Natl Acad Sci U S A* **102:** 10682-10687.

Hutchison, C.A., 3rd. 2007. DNA sequencing: bench to bedside and beyond. *Nucleic Acids Res* **35:** 6227-6237.

Hutt, F.B. 1933. Genetics of the Fowl. II. a Four-Gene Autosomal Linkage Group. *Genetics* **18:** 82–94.

Hutt, F.B. 1936. Genetics of the fowl. VI. A tentative chromosome map. *Neue Forsch in Tierzucht u Abst (Duerst Festschrift)***:** 105–112.

Ianella, P., L.P. Venancio, N.B. Stafuzza, M.N. Miziara, R. Agarwala, A.A. Schaffer, P.K. Riggs, J.E. Womack, and M.E. Amaral. 2008. First radiation hybrid map of the river buffalo X chromosome (BBUX) and comparison with BTAX. *Anim Genet* **39:** 196-200.

ICGC. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432:** 695-716.

Itoh, T., T. Watanabe, N. Ihara, P. Mariani, C.W. Beattie, Y. Sugimoto, and A. Takasuga. 2005. A comprehensive radiation hybrid map of the bovine genome comprising 5593 loci. *Genomics* **85:** 413-424.

James, L.C., A.H. Keeble, Z. Khan, D.A. Rhodes, and J. Trowsdale. 2007. Structural basis for PRYSPRY-mediated tripartite motif (TRIM) protein function. *Proc Natl Acad Sci U S A* **104:** 6200-6205.

Jarvis, E.D. 2004. Learned birdsong and the neurobiology of human language. *Ann N Y Acad Sci* **1016:** 749-777.

Jennen, D.G., R.P. Crooijmans, M. Morisson, A.E. Grootemaat, J.J. Van Der Poel, A. Vignal, and M.A. Groenen. 2004. A radiation hybrid map of chicken chromosome 15. *Anim Genet* **35:** 63-65.

Jeong, J., A.U. Rao, J. Xu, S.L. Ogg, Y. Hathout, C. Fenselau, and I.H. Mather. 2009. The PRY/SPRY/B30.2 domain of butyrophilin 1A1 (BTN1A1) binds to xanthine oxidoreductase: implications for the function of BTN1A1 in the mammary gland and other tissues. *J Biol Chem* **284:** 22444-22456.

Jones, H.B. 1996. Hybrid selection as a method of increasing mapping power for radiation hybrids. *Genome Res* **6:** 761-769.

Ju, J., D.H. Kim, L. Bi, Q. Meng, X. Bai, Z. Li, X. Li, M.S. Marma, S. Shi, J. Wu, J.R. Edwards, A. Romu, and N.J. Turro. 2006. Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators. *Proc Natl Acad Sci U S A* **103:** 19635-19640.

Kao, F.T. 1973. Identification of chick chromosomes in cell hybrids formed between chick erythrocytes and adenine-requiring mutants of Chinese hamster cells. *Proc Natl Acad Sci U S A* **70:** 2893-2898.

Karere, G.M., L.A. Lyons, and L. Froenicke. 2010. Enhancing radiation hybrid mapping through whole genome amplification. *Hereditas* **147:** 103-112.

Kato, H., O. Takeuchi, S. Sato, M. Yoneyama, M. Yamamoto, K. Matsui, S. Uematsu, A. Jung, T. Kawai, K.J. Ishii, O. Yamaguchi, K. Otsu, T. Tsujimura, C.S. Koh, C. Reis e Sousa, Y. Matsuura, T. Fujita, and S. Akira. 2006. Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses. *Nature* **441:** 101-105.

Kayang, B.B., V. Fillon, M. Inoue-Murayama, M. Miwa, S. Leroux, K. Feve, J.L. Monvoisin, F. Pitel, M. Vignoles, C. Mouilhayrat, C. Beaumont, S. Ito, F. Minvielle, and A. Vignal. 2006. Integrated maps in quail (Coturnix japonica) confirm the high degree of synteny conservation with chicken (Gallus gallus) despite 35 million years of divergence. *BMC Genomics* **7:** 101.

Kida, H., R. Yanagawa, and Y. Matsuoka. 1980. Duck influenza lacking evidence of disease signs and immune response. *Infect Immun* **30:** 547-553.

Kiguwa, S.L., P. Hextall, A.L. Smith, R. Critcher, J. Swinburne, L. Millon, M.M. Binns, P.N. Goodfellow, L.C. McCarthy, C.J. Farr, and E.A. Oakenfull. 2000. A horse whole-genome-radiation hybrid panel: chromosome 1 and 10 preliminary maps. *Mamm Genome* **11:** 803-805.

Kim, J.K., N.J. Negovetich, H.L. Forrest, and R.G. Webster. 2009. Ducks: the "Trojan horses" of H5N1 influenza. *Influenza Other Respi Viruses* **3:** 121-128.

Korlach, J., A. Bibillo, J. Wegener, P. Peluso, T.T. Pham, I. Park, S. Clark, G.A. Otto, and S.W. Turner. 2008. Long, processive enzymatic DNA synthesis using 100% dye-labeled terminal phosphate-linked nucleotides. *Nucleosides Nucleotides Nucleic Acids* **27:** 1072-1083.

Korlach, J., K.P. Bjornson, B.P. Chaudhuri, R.L. Cicero, B.A. Flusberg, J.J. Gray, D. Holden, R. Saxena, J. Wegener, and S.W. Turner. 2010. Real-time DNA sequencing from single polymerase molecules. *Methods Enzymol* **472:** 431-455.

Kramerov, D.A. and N.S. Vassetzky. 2005. Short retroposons in eukaryotic genomes. *Int Rev Cytol* **247:** 165-221.

Kraus, R.H., H.H. Kerstens, P. Van Hooft, R.P. Crooijmans, J.J. Van Der Poel, J. Elmberg, A. Vignal, Y. Huang, N. Li, H.H. Prins, and M.A. Groenen. 2011. Genome wide SNP discovery, analysis and evaluation in mallard (Anas platyrhynchos). *BMC Genomics* **12:** 150.

Kuhl, P.K. 2003. Human speech and birdsong: communication and the social brain. *Proc Natl Acad Sci U S A* **100:** 9645-9646.

Kumar, S., A. Sood, J. Wegener, P.J. Finn, S. Nampalli, J.R. Nelson, A. Sekher, P. Mitsis, J. Macklin, and C.W. Fuller. 2005. Terminal phosphate labeled nucleotides: synthesis, applications, and linker effect on incorporation by DNA polymerases. *Nucleosides Nucleotides Nucleic Acids* **24:** 401-408.

Kuroda, Y., N. Arai, M. Arita, M. Teranishi, T. Hori, M. Harata, and S. Mizuno. 2001. Absence of Z-chromosome inactivation for five genes in male chickens. *Chromosome Res* **9:** 457-468.

Kuroiwa, A., T. Yokomine, H. Sasaki, M. Tsudzuki, K. Tanaka, T. Namikawa, and Y. Matsuda. 2002. Biallelic expression of Z-linked genes in male chickens. *Cytogenet Genome Res* **99:** 310-314.

Kwok, C., R. Critcher, and K. Schmitt. 1999. Construction and characterization of zebrafish whole genome radiation hybrids. *Methods Cell Biol* **60:** 287-302.

Kwok, C., R.M. Korn, M.E. Davis, D.W. Burt, R. Critcher, L. McCarthy, B.H. Paw, L.I. Zon, P.N. Goodfellow, and K. Schmitt. 1998. Characterization of whole genome radiation hybrid mapping resources for non-mammalian vertebrates. *Nucleic Acids Res* **26:** 3562-3566.

Ladjali-Mohammededi, K., J.J. Bitgood, M. Tixier-Boichard, and F.A. Ponce de Leon. 1999. International system for standarized avian karyotype (ISSAK): standardized banded karyotypes of the domestic fowl (*Gallus domesticus*). *Cytogenet Cell Genet***:** 271-276.

Lam, W.L., T.S. Lee, and W. Gilbert. 1996. Active transposition in zebrafish. *Proc Natl Acad Sci U S A* **93:** 10870-10875.

Lander, E.S. L.M. Linton B. Birren C. Nusbaum M.C. Zody J. Baldwin K. Devon K. Dewar M. Doyle W. FitzHugh R. Funke D. Gage K. Harris A. Heaford J. Howland L. Kann J. Lehoczky R. LeVine P. McEwan K. McKernan J. Meldrim J.P. Mesirov C. Miranda W. Morris J. Naylor C. Raymond M. Rosetti R. Santos A. Sheridan C. Sougnez N. Stange-Thomann N. Stojanovic A. Subramanian D. Wyman J. Rogers J. Sulston R. Ainscough S. Beck D. Bentley J. Burton C. Clee N. Carter A. Coulson R. Deadman P. Deloukas A. Dunham I. Dunham R. Durbin L. French D. Grafham S. Gregory T. Hubbard S. Humphray A. Hunt M. Jones C. Lloyd A. McMurray L. Matthews S. Mercer S. Milne J.C. Mullikin A. Mungall R. Plumb M. Ross R. Shownkeen S. Sims R.H. Waterston R.K. Wilson L.W. Hillier J.D. McPherson M.A. Marra E.R. Mardis

L.A. Fulton A.T. Chinwalla K.H. Pepin W.R. Gish S.L. Chissoe M.C. Wendl K.D. Delehaunty T.L. Miner A. Delehaunty J.B. Kramer L.L. Cook R.S. Fulton D.L. Johnson P.J. Minx S.W. Clifton T. Hawkins E. Branscomb P. Predki P. Richardson S. Wenning T. Slezak N. Doggett J.F. Cheng A. Olsen S. Lucas C. Elkin E. Uberbacher M. Frazier R.A. Gibbs D.M. Muzny S.E. Scherer J.B. Bouck E.J. Sodergren K.C. Worley C.M. Rives J.H. Gorrell M.L. Metzker S.L. Naylor R.S. Kucherlapati D.L. Nelson G.M. Weinstock Y. Sakaki A. Fujiyama M. Hattori T. Yada A. Toyoda T. Itoh C. Kawagoe H. Watanabe Y. Totoki T. Taylor J. Weissenbach R. Heilig W. Saurin F. Artiguenave P. Brottier T. Bruls E. Pelletier C. Robert P. Wincker D.R. Smith L. Doucette-Stamm M. Rubenfield K. Weinstock H.M. Lee J. Dubois A. Rosenthal M. Platzer G. Nyakatura S. Taudien A. Rump H. Yang J. Yu J. Wang G. Huang J. Gu L. Hood L. Rowen A. Madan S. Qin R.W. Davis N.A. Federspiel A.P. Abola M.J. Proctor R.M. Myers J. Schmutz M. Dickson J. Grimwood D.R. Cox M.V. Olson R. Kaul N. Shimizu K. Kawasaki S. Minoshima G.A. Evans M. Athanasiou R. Schultz B.A. Roe F. Chen H. Pan J. Ramser H. Lehrach R. Reinhardt W.R. McCombie M. de la Bastide N. Dedhia H. Blocker K. Hornischer G. Nordsiek R. Agarwala L. Aravind J.A. Bailey A. Bateman S. Batzoglou E. Birney P. Bork D.G. Brown C.B. Burge L. Cerutti H.C. Chen D. Church M. Clamp R.R. Copley T. Doerks S.R. Eddy E.E. Eichler T.S. Furey J. Galagan J.G. Gilbert C. Harmon Y. Hayashizaki D. Haussler H. Hermjakob K. Hokamp W. Jang L.S. Johnson T.A. Jones S. Kasif A. Kaspryzk S. Kennedy W.J. Kent P. Kitts E.V. Koonin I. Korf D. Kulp D. Lancet T.M. Lowe A. McLysaght T. Mikkelsen J.V. Moran N. Mulder V.J. Pollara C.P. Ponting G. Schuler J. Schultz G. Slater A.F. Smit E. Stupka J. Szustakowski D. Thierry-Mieg J. Thierry-Mieg L. Wagner J. Wallis R. Wheeler A. Williams Y.I. Wolf K.H. Wolfe S.P. Yang R.F. Yeh F. Collins M.S. Guyer J. Peterson A. Felsenfeld K.A. Wetterstrand A. Patrinos M.J. Morgan P. de Jong J.J. Catanese K. Osoegawa H. Shizuya S. Choi and Y.J. Chen. 2001. Initial sequencing and analysis of the human genome. *Nature* **409:** 860-921.

Lasken, R.S. 2009. Genomic DNA amplification by the multiple displacement amplification (MDA) method. *Biochem Soc Trans* **37:** 450-453.

Laurent, P., L. Schibler, A. Vaiman, J. Laubier, C. Delcros, G. Cosseddu, D. Vaiman, E.P. Cribiu, and M. Yerle. 2007. A 12 000-rad whole-genome radiation hybrid panel in sheep: application to the study of the ovine chromosome 18 region containing a QTL for scrapie susceptibility. *Anim Genet* **38:** 358-363.

Leroux, S., M. Dottax, S. Bardes, F. Vignoles, K. Feve, F. Pitel, M. Morisson, and A. Vignal. 2005. Construction of a radiation hybrid map of chicken chromosome 2 and alignment to the chicken draft sequence. *BMC Genomics* **6:** 12.

Levene, M.J., J. Korlach, S.W. Turner, M. Foquet, H.G. Craighead, and W.W. Webb. 2003. Zero-mode waveguides for single-molecule analysis at high concentrations. *Science* **299:** 682-686.

Lewin, H.A., D.M. Larkin, J. Pontius, and S.J. O'Brien. 2009. Every genome sequence needs a good map. *Genome Res* **19:** 1925-1928.

Li, M., Y. Hou, J. Wang, X. Chen, Z.M. Shao, and X.M. Yin. 2011a. Kinetics comparisons of mammalian Atg4 homologues indicate selective preferences toward diverse Atg8 substrates. *J Biol Chem* **286:** 7327-7338.

Li, Q., N. Li, X. Hu, J. Li, Z. Du, L. Chen, G. Yin, J. Duan, H. Zhang, Y. Zhao, and J. Wang. 2011b. Genome-wide mapping of DNA methylation in chicken. *PLoS One* **6:** e19428.

Li, R. W. Fan G. Tian H. Zhu L. He J. Cai Q. Huang Q. Cai B. Li Y. Bai Z. Zhang Y. Zhang W. Wang J. Li F. Wei H. Li M. Jian R. Nielsen D. Li W. Gu Z. Yang Z. Xuan O.A. Ryder F.C. Leung Y. Zhou J. Cao X. Sun Y. Fu X. Fang X. Guo B. Wang R. Hou F. Shen B. Mu P. Ni R. Lin W. Qian G. Wang C. Yu W. Nie J. Wang Z. Wu H. Liang J. Min Q. Wu S. Cheng J. Ruan M. Wang Z. Shi M. Wen B. Liu X. Ren H. Zheng D. Dong K. Cook G. Shan H. Zhang C. Kosiol X. Xie Z. Lu Y. Li C.C. Steiner T.T. Lam S. Lin Q. Zhang G. Li J. Tian T. Gong H. Liu D. Zhang L. Fang C. Ye J. Zhang W. Hu A. Xu Y. Ren G. Zhang M.W. Bruford Q. Li L. Ma Y. Guo N. An Y. Hu Y. Zheng Y. Shi Z. Li Q. Liu Y. Chen J. Zhao N. Qu S. Zhao F. Tian X. Wang H. Wang L. Xu X. Liu T. Vinar Y. Wang T.W. Lam S.M. Yiu S. Liu Y. Huang G. Yang Z. Jiang N. Qin L. Li L. Bolund K. Kristiansen G.K. Wong M. Olson X. Zhang S. Li and H. Yang. 2010a. The sequence and de novo assembly of the giant panda genome. *Nature* **463:** 311-317.

Li, R., H. Zhu, J. Ruan, W. Qian, X. Fang, Z. Shi, Y. Li, S. Li, G. Shan, K. Kristiansen, H. Yang, and J. Wang. 2010b. De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res* **20:** 265-272.

Lovmar, L. and A.C. Syvanen. 2006. Multiple displacement amplification to create a long-lasting source of DNA for genetic studies. *Hum Mutat* **27:** 603-614.

Lunetta, K.L., M. Boehnke, K. Lange, and D.R. Cox. 1995. Experimental design and error detection for polyploid radiation hybrid mapping. *Genome Res* **5:** 151-163.

MacDonald, M.R., S.M. Veniamin, X. Guo, J. Xia, D.A. Moon, and K.E. Magor. 2007. Genomics of antiviral defenses in the duck, a natural host of influenza and hepatitis B viruses. *Cytogenet Genome Res* **117:** 195-206.

Malone, J.H. and B. Oliver. 2008. The sex chromosome that refused to die. *Bioessays* **30:** 409-411.

Mardis, E.R. 2008. Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet* **9:** 387-402.

Margulies, M., M. Egholm, W.E. Altman, S. Attiya, J.S. Bader, L.A. Bemben, J. Berka, M.S. Braverman, Y.J. Chen, Z. Chen, S.B. Dewell, L. Du, J.M. Fierro, X.V. Gomes, B.C.

Godwin, W. He, S. Helgesen, C.H. Ho, G.P. Irzyk, S.C. Jando, M.L. Alenquer, T.P. Jarvie, K.B. Jirage, J.B. Kim, J.R. Knight, J.R. Lanza, J.H. Leamon, S.M. Lefkowitz, M. Lei, J. Li, K.L. Lohman, H. Lu, V.B. Makhijani, K.E. McDade, M.P. McKenna, E.W. Myers, E. Nickerson, J.R. Nobile, R. Plant, B.P. Puc, M.T. Ronan, G.T. Roth, G.J. Sarkis, J.F. Simons, J.W. Simpson, M. Srinivasan, K.R. Tartaro, A. Tomasz, K.A. Vogt, G.A. Volkmer, S.H. Wang, Y. Wang, M.P. Weiner, P. Yu, R.F. Begley, and J.M. Rothberg. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437:** 376-380.

Marie-Etancelin, C., Chapuis H., Brun J.M., Marzul C., Mialon-Richard M.M., and R. R. 2008. Genetics and selection of mule ducks in France: a review. *Worlds Poult Sci J***:** 187-207.

Marques, E., S. de Givry, P. Stothard, B. Murdoch, Z. Wang, J. Womack, and S.S. Moore. 2007. A high resolution radiation hybrid map of bovine chromosome 14 identifies scaffold rearrangement in the latest bovine assembly. *BMC Genomics* **8:** 254.

Martin, S.L., W.L. Li, A.V. Furano, and S. Boissinot. 2005. The structures of mouse and human L1 elements reflect their insertion mechanism. *Cytogenet Genome Res* **110:** 223-228.

Mathias, S.L., A.F. Scott, H.H. Kazazian, Jr., J.D. Boeke, and A. Gabriel. 1991. Reverse transcriptase encoded by a human transposable element. *Science* **254:** 1808-1810.

Maxam, A.M. and W. Gilbert. 1977. A new method for sequencing DNA. *Proc Natl Acad Sci U S A* **74:** 560-564.

McCarthy, L.C., M.T. Bihoreau, S.L. Kiguwa, J. Browne, T.K. Watanabe, H. Hishigaki, A. Tsuji, S. Kiel, C. Webber, M.E. Davis, C. Knights, A. Smith, R. Critcher, P. Huxtall, J.R. Hudson, Jr., T. Ono, H. Hayashi, T. Takagi, Y. Nakamura, A. Tanigami, P.N. Goodfellow, G.M. Lathrop, and M.R. James. 2000. A whole-genome radiation hybrid panel and framework map of the rat genome. *Mamm Genome* **11:** 791-795.

McCarthy, L.C., J. Terrett, M.E. Davis, C.J. Knights, A.L. Smith, R. Critcher, K. Schmitt, J. Hudson, N.K. Spurr, and P.N. Goodfellow. 1997. A first-generation whole genome-radiation hybrid map spanning the mouse genome. *Genome Res* **7:** 1153-1161.

McQueen, H.A., J. Fantes, S.H. Cross, V.H. Clark, A.L. Archibald, and A.P. Bird. 1996. CpG islands of chicken are concentrated on microchromosomes. *Nat Genet* **12:** 321-324.

McQueen, H.A., D. McBride, G. Miele, A.P. Bird, and M. Clinton. 2001. Dosage compensation in birds. *Curr Biol* **11:** 253-257.

McQueen, H.A., G. Siriaco, and A.P. Bird. 1998. Chicken microchromosomes are hyperacetylated, early replicating, and gene rich. *Genome Res* **8:** 621-630.

Melamed, E. and A.P. Arnold. 2007. Regional differences in dosage compensation on the chicken Z chromosome. *Genome Biol* **8:** R202.

Migeon, B.R. and C.S. Miller. 1968. Human-mouse somatic cell hybrids with single human chromosome (group E): link with thymidine kinase activity. *Science* **162:** 1005-1006.

Minna, J.D. and H.G. Coon. 1974. Human times mouse hybrid cells segregating mouse chromosomes and isozymes. *Nature* **252:** 401-404.

Minna, J.D., A.F. Gazdar, G.M. Iverson, T.H. Marshall, K. Stormberg, and S.H. Wilson. 1974. Oncornavirus expression in human x mouse hybrid cells segregating mouse chromosomes. *Proc Natl Acad Sci U S A* **71:** 1695-1700.

Mitra, R.D., J. Shendure, J. Olejnik, O. Edyta Krzymanska, and G.M. Church. 2003. Fluorescent in situ sequencing on polymerase colonies. *Anal Biochem* **320:** 55-65.

Moon, D.A. and K.E. Magor. 2004. Construction and characterization of a fosmid library for comparative analysis of the duck genome. *Anim Genet* **35:** 417-418.

Morisson, M., M. Denis, D. Milan, C. Klopp, S. Leroux, S. Bardes, F. Pitel, F. Vignoles, M. Gerus, V. Fillon, M. Douaud, and A. Vignal. 2007. The chicken RH map: current state of progress and microchromosome mapping. *Cytogenet Genome Res* **117:** 14-21.

Morisson, M., C. Jiguet-Jiglaire, S. Leroux, T. Faraut, S. Bardes, K. Feve, C. Genet, F. Pitel, D. Milan, and A. Vignal. 2004. Development of a gene-based radiation hybrid map of chicken Chromosome 7 and comparison to human and mouse. *Mamm Genome* **15:** 732-739.

Morisson, M., A. Lemiere, S. Bosc, M. Galan, F. Plisson-Petit, P. Pinton, C. Delcros, K. Feve, F. Pitel, V. Fillon, M. Yerle, and A. Vignal. 2002. ChickRH6: a chicken whole-genome radiation hybrid panel. *Genet Sel Evol* **34:** 521-533.

Morisson, M., S. Leroux, C. Jiguet-Jiglaire, S. Assaf, F. Pitel, S. Lagarrigue, S. Bardes, K. Feve, T. Faraut, D. Milan, and A. Vignal. 2005. A gene-based radiation hybrid map of chicken microchromosome 14: comparison to human and alignment to the assembled chicken sequence. *Genet Sel Evol* **37:** 229-251.

Mortazavi, A., B.A. Williams, K. McCue, L. Schaeffer, and B. Wold. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5:** 621-628.

Muller, H.J. 1964. The Relation of Recombination to Mutational Advance. *Mutat Res* **106:** 2-9.

Muller, J. and R.R. Reisz. 2005. Four well-constrained calibration points from the vertebrate fossil record for molecular clock estimates. *Bioessays* **27:** 1069-1075.

Murphy, W.J., M. Menotti-Raymond, L.A. Lyons, M.A. Thompson, and S.J. O'Brien. 1999. Development of a feline whole genome radiation hybrid panel and comparative mapping of human chromosome 12 and 22 loci. *Genomics* **57:** 1-8.

Murphy, W.J., J.E. Page, C. Smith, Jr., R.C. Desrosiers, and S.J. O'Brien. 2001. A radiation hybrid mapping panel for the rhesus macaque. *J Hered* **92:** 516-519.

Nabholz, M., V. Miggiano, and W. Bodmer. 1969. Genetic analysis with human--mouse somatic cell hybrids. *Nature* **223:** 358-363.

Nanda, I., P. Benisch, D. Fetting, T. Haaf, and M. Schmid. 2011. Synteny conservation of chicken macrochromosomes 1-10 in different avian lineages revealed by cross-species chromosome painting. *Cytogenet Genome Res* **132:** 165-181.

Nanda, I. and M. Schmid. 1994. Localization of the telomeric (TTAGGG)n sequence in chicken (*Gallus domesticus*) chromosomes. *Cytogenet Cell Genet***:** 190-193.

Nanda, I., Z. Shan, M. Schartl, D.W. Burt, M. Koehler, H. Nothwang, F. Grutzner, I.R. Paton, D. Windsor, I. Dunn, W. Engel, P. Staeheli, S. Mizuno, T. Haaf, and M. Schmid. 1999. 300 million years of conserved synteny between chicken Z and human chromosome 9. *Nat Genet* **21:** 258-259.

Nguyen, T., X.K. Liu, Y. Zhang, and C. Dong. 2006. BTNL2, a butyrophilin-like molecule that functions to inhibit T cell activation. *J Immunol* **176:** 7354-7360.

Okazaki, K., A. Takada, T. Ito, M. Imai, H. Takakuwa, M. Hatta, H. Ozaki, T. Tanizaki, T. Nagano, A. Ninomiya, V.A. Demenev, M.M. Tyaptirganov, T.D. Karatayeva, S.S. Yamnikova, D.K. Lvov, and H. Kida. 2000. Precursor genes of future pandemic influenza viruses are perpetuated in ducks nesting in Siberia. *Arch Virol* **145:** 885-893.

Olivier, M., A. Aggarwal, J. Allen, A.A. Almendras, E.S. Bajorek, E.M. Beasley, S.D. Brady, J.M. Bushard, V.I. Bustos, A. Chu, T.R. Chung, A. De Witte, M.E. Denys, R. Dominguez, N.Y. Fang, B.D. Foster, R.W. Freudenberg, D. Hadley, L.R. Hamilton, T.J. Jeffrey, L. Kelly, L. Lazzeroni, M.R. Levy, S.C. Lewis, X. Liu, F.J. Lopez, B. Louie, J.P. Marquis, R.A. Martinez, M.K. Matsuura, N.S. Misherghi, J.A. Norton, A. Olshen, S.M. Perkins, A.J. Perou, C. Piercy, M. Piercy, F. Qin, T. Reif, K. Sheppard, V. Shokoohi, G.A. Smick, W.L. Sun, E.A. Stewart, J. Fernando, Tejeda, N.M. Tran, T. Trejo, N.T. Vo, S.C. Yan, D.L. Zierten, S. Zhao, R. Sachidanandam, B.J. Trask, R.M. Myers, and D.R. Cox. 2001. A high-resolution radiation hybrid map of the human genome draft sequence. *Science* **291:** 1298-1302.

Organ, C.L., A.M. Shedlock, A. Meade, M. Pagel, and S.V. Edwards. 2007. Origin of avian genome size and structure in non-avian dinosaurs. *Nature* **446:** 180-184.

Otto, S.P., J.R. Pannell, C.L. Peichel, T.L. Ashman, D. Charlesworth, A.K. Chippindale, L.F. Delph, R.F. Guerrero, S.V. Scarpino, and B.F. McAllister. 2011. About PAR: the distinct evolutionary dynamics of the pseudoautosomal region. *Trends Genet* **27:** 358-367.

Ozsolak, F., A.R. Platt, D.R. Jones, J.G. Reifenberger, L.E. Sass, P. McInerney, J.F. Thompson, J. Bowers, M. Jarosz, and P.M. Milos. 2009. Direct RNA sequencing. *Nature* **461:** 814-818.

Page, J.E. and W.J. Murphy. 2008. Construction of radiation hybrid panels. *Methods Mol Biol* **422:** 51-64.

Pareek, C.S., R. Smoczynski, and A. Tretyn. 2011. Sequencing technologies and genome sequencing. *J Appl Genet* **52:** 413-435.

Pask, R., H.E. Rance, B.J. Barratt, S. Nutland, D.J. Smyth, M. Sebastian, R.C. Twells, A. Smith, A.C. Lam, L.J. Smink, N.M. Walker, and J.A. Todd. 2004. Investigating the utility of combining phi29 whole genome amplification and highly multiplexed single nucleotide polymorphism BeadArray genotyping. *BMC Biotechnol* **4:** 15.

Peiris, J.S., M.D. de Jong, and Y. Guan. 2007. Avian influenza virus (H5N1): a threat to human health. *Clin Microbiol Rev* **20:** 243-267.

Pereira, S.L. and A.J. Baker. 2006. A mitogenomic timescale for birds detects variable phylogenetic rates of molecular evolution and refutes the standard molecular clock. *Mol Biol Evol* **23:** 1731-1740.

Pitel, F., B. Abasht, M. Morisson, R.P. Crooijmans, F. Vignoles, S. Leroux, K. Feve, S. Bardes, D. Milan, S. Lagarrigue, M.A. Groenen, M. Douaire, and A. Vignal. 2004. A high-resolution radiation hybrid map of chicken chromosome 5 and comparison with human chromosomes. *BMC Genomics* **5:** 66.

Ponce de Leon, F.A., Y. Li, and Z. Weng. 1992. Early and late replicative chromosomal banding patterns of Gallus domesticus. *J Hered* **83:** 36-42.

Prasad, A., T. Schiex, S. McKay, B. Murdoch, Z. Wang, J.E. Womack, P. Stothard, and S.S. Moore. 2007. High resolution radiation hybrid maps of bovine chromosomes 19 and 29: comparison with the bovine genome sequence assembly. *BMC Genomics* **8:** 310.

Prestel, M., C. Feller, T. Straub, H. Mitlohner, and P.B. Becker. 2010. The activation potential of MOF is constrained for dosage compensation. *Mol Cell* **38:** 815-826.

Rabie, T.S., R.P. Crooijmans, M. Morisson, J. Andryszkiewicz, J.J. van der Poel, A. Vignal, and M.A. Groenen. 2004. A radiation hybrid map of chicken Chromosome 4. *Mamm Genome* **15:** 560-569.

Rao, M., M. Morisson, T. Faraut, S. Bardes, K. Feve, E. Labarthe, Y. Huang, N. Li, and A. Vignal. 2012. A duck RH panel and its potential for assisting NGS genome assembly.

Reed, K.M., L.D. Chaves, M.K. Hall, T.P. Knutson, and D.E. Harry. 2005. A comparative genetic map of the turkey genome. *Cytogenet Genome Res* **111:** 118-127.

Reed, K.M., L.D. Chaves, and K.M. Mendoza. 2007. An integrated and comparative genetic map of the turkey genome. *Cytogenet Genome Res* **119:** 113-126.

Rexroad, C.E., 3rd, E.K. Owens, J.S. Johnson, and J.E. Womack. 2000. A 12,000 rad whole genome radiation hybrid panel for high resolution mapping in cattle: characterization of the centromeric end of chromosome 1. *Anim Genet* **31:** 262-265.

Rodionov, A.V. 1996. [Micro vs. macro: structural-functional organization of avian micro- and macrochromosomes]. *Genetika* **32:** 597-608.

Ronaghi, M., S. Karamohamed, B. Pettersson, M. Uhlen, and P. Nyren. 1996. Real-time DNA sequencing using detection of pyrophosphate release. *Anal Biochem* **242:** 84-89.

Rothberg, J.M., W. Hinz, T.M. Rearick, J. Schultz, W. Mileski, M. Davey, J.H. Leamon, K. Johnson, M.J. Milgrew, M. Edwards, J. Hoon, J.F. Simons, D. Marran, J.W. Myers, J.F. Davidson, A. Branting, J.R. Nobile, B.P. Puc, D. Light, T.A. Clark, M. Huber, J.T. Branciforte, I.B. Stoner, S.E. Cawley, M. Lyons, Y. Fu, N. Homer, M. Sedova, X. Miao, B. Reed, J. Sabina, E. Feierstein, M. Schorn, M. Alanjary, E. Dimalanta, D. Dressman, R. Kasinskas, T. Sokolsky, J.A. Fidanza, E. Namsaraev, K.J. McKernan, A.

Williams, G.T. Roth, and J. Bustillo. 2011. An integrated semiconductor device enabling non-optical genome sequencing. *Nature* **475:** 348-352.

Rubin, C.J., M.C. Zody, J. Eriksson, J.R. Meadows, E. Sherwood, M.T. Webster, L. Jiang, M. Ingman, T. Sharpe, S. Ka, F. Hallbook, F. Besnier, O. Carlborg, B. Bed'hom, M. Tixier-Boichard, P. Jensen, P. Siegel, K. Lindblad-Toh, and L. Andersson. 2010. Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* **464:** 587-591.

Ruddle, F.H. 1973. Linkage analysis in man by somatic cell genetics. *Nature* **242:** 165-169.

Sanger, F., G.M. Air, B.G. Barrell, N.L. Brown, A.R. Coulson, C.A. Fiddes, C.A. Hutchison, P.M. Slocombe, and M. Smith. 1977. Nucleotide sequence of bacteriophage phi X174 DNA. *Nature* **265:** 687-695.

Sanger, F. and A.R. Coulson. 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol* **94:** 441-448.

Sarropoulou, E., R. Franch, B. Louro, D.M. Power, L. Bargelloni, A. Magoulas, F. Senger, M. Tsalavouta, T. Patarnello, F. Galibert, G. Kotoulas, and R. Geisler. 2007. A gene-based radiation hybrid map of the gilthead sea bream Sparus aurata refines and exploits conserved synteny with Tetraodon nigroviridis. *BMC Genomics* **8:** 44.

Sasazaki, S., T. Hinenoya, B. Lin, A. Fujiwara, and H. Mannen. 2006. A comparative map of macrochromosomes between chicken and Japanese quail based on orthologous genes. *Anim Genet* **37:** 316-320.

Schadt, E.E., S. Turner, and A. Kasarskis. 2010. A window into third-generation sequencing. *Hum Mol Genet* **19:** R227-240.

Scherz-Shouval, R., Y. Sagiv, H. Shorer, and Z. Elazar. 2003. The COOH terminus of GATE-16, an intra-Golgi transport modulator, is cleaved by the human cysteine protease HsApg4A. *J Biol Chem* **278:** 14053-14058.

Scherz-Shouval, R., E. Shvets, E. Fass, H. Shorer, L. Gil, and Z. Elazar. 2007. Reactive oxygen species are essential for autophagy and specifically regulate the activity of Atg4. *Embo J* **26:** 1749-1760.

Schimke, R.T. 1984. Gene amplification in cultured animal cells. *Cell* **37:** 705-713.

Schmid, M., E. Enderle, D. Schindler, and W. Schempp. 1989. Chromosome banding and DNA replication patterns in bird karyotypes. *Cytogenet Cell Genet* **52:** 139-146.

Schmitt, K., J.W. Foster, R.W. Feakes, C. Knights, M.E. Davis, D.J. Spillett, and P.N. Goodfellow. 1996. Construction of a mouse whole-genome radiation hybrid panel and application to MMU11. *Genomics* **34:** 193-197.

Scholz, B., K. Kultima, A. Mattsson, J. Axelsson, B. Brunstrom, K. Halldin, M. Stigson, and L. Dencker. 2006. Sex-dependent gene expression in early brain development of chicken embryos. *BMC Neurosci* **7:** 12.

Senger, F., C. Priat, C. Hitte, E. Sarropoulou, R. Franch, R. Geisler, L. Bargelloni, D. Power, and F. Galibert. 2006. The first radiation hybrid map of a perch-like fish: the gilthead seabream (Sparus aurata L). *Genomics* **87:** 793-800.

Shendure, J. and H. Ji. 2008. Next-generation DNA sequencing. *Nat Biotechnol* **26:** 1135-1145.

Shendure, J., G.J. Porreca, N.B. Reppas, X. Lin, J.P. McCutcheon, A.M. Rosenbaum, M.D. Wang, K. Zhang, R.D. Mitra, and G.M. Church. 2005. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* **309:** 1728-1732.

Shibusawa, M., M. Nishibori, C. Nishida-Umehara, M. Tsudzuki, J. Masabanda, D.K. Griffin, and Y. Matsuda. 2004. Karyotypic evolution in the Galliformes: an examination of the process of karyotypic evolution by comparison of the molecular cytogenetic findings with the molecular phylogeny. *Cytogenet Genome Res* **106:** 111-119.

Shirasu, K., A.H. Schulman, T. Lahaye, and P. Schulze-Lefert. 2000. A contiguous 66-kb barley DNA sequence provides evidence for reversible genome expansion. *Genome Res* **10:** 908-915.

Siden, T.S., J. Kumlien, C.E. Schwartz, and D. Rohme. 1992. Radiation fusion hybrids for human chromosomes 3 and X generated at various irradiation doses. *Somat Cell Mol Genet* **18:** 33-44.

Silander, K. and J. Saarela. 2008. Whole genome amplification with Phi29 DNA polymerase to enable genetic or genomic analysis of samples of low DNA yield. *Methods Mol Biol* **439:** 1-18.

Sjoblom, T., S. Jones, L.D. Wood, D.W. Parsons, J. Lin, T.D. Barber, D. Mandelker, R.J. Leary, J. Ptak, N. Silliman, S. Szabo, P. Buckhaults, C. Farrell, P. Meeh, S.D. Markowitz, J. Willis, D. Dawson, J.K. Willson, A.F. Gazdar, J. Hartigan, L. Wu, C. Liu, G. Parmigiani, B.H. Park, K.E. Bachman, N. Papadopoulos, B. Vogelstein, K.W. Kinzler, and V.E. Velculescu. 2006. The consensus coding sequences of human breast and colorectal cancers. *Science* **314:** 268-274.

Skinner, B.M. and D.K. Griffin. 2011. Intrachromosomal rearrangements in avian genome evolution: evidence for regions prone to breakpoints. *Heredity (Edinb)* **108:** 37-41.

Skinner, B.M., L.B. Robertson, H.G. Tempest, E.J. Langley, D. Ioannou, K.E. Fowler, R.P. Crooijmans, A.D. Hall, D.K. Griffin, and M. Volker. 2009. Comparative genomics in chicken and Pekin duck using FISH mapping and microarray analysis. *BMC Genomics* **10:** 357.

Smit, A.F., G. Toth, A.D. Riggs, and J. Jurka. 1995. Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences. *J Mol Biol* **246:** 401-417.

Smith, J., C.K. Bruley, I.R. Paton, I. Dunn, C.T. Jones, D. Windsor, D.R. Morrice, A.S. Law, J. Masabanda, A. Sazanov, D. Waddington, R. Fries, and D.W. Burt. 2000. Differences in gene density on chicken macrochromosomes and microchromosomes. *Anim Genet* **31:** 96-103.

Snapir, N., J. Rulf, A. Meltzer, G. Gvaryahu, I. Rozenboim, and B. Robinzon. 1998. Testosterone concentrations, testes weight and morphology of mule drakes (Muscovy drake x Khaki Campbell). *Br Poult Sci* **39:** 572-574.

Songserm, T., R. Jam-on, N. Sae-Heng, N. Meemak, D.J. Hulse-Post, K.M. Sturm-Ramirez, and R.G. Webster. 2006. Domestic ducks and H5N1 influenza epidemic, Thailand. *Emerg Infect Dis* **12:** 575-581.

Springer, M.S., W.J. Murphy, E. Eizirik, and S.J. O'Brien. 2003. Placental mammal diversification and the Cretaceous-Tertiary boundary. *Proc Natl Acad Sci U S A* **100:** 1056-1061.

Spurgeon, S.L., R.C. Jones, and R. Ramakrishnan. 2008. High throughput gene expression measurement with real time PCR in a microfluidic dynamic array. *PLoS One* **3:** e1662.

Stapley, J., T.R. Birkhead, T. Burke, and J. Slate. 2008. A linkage map of the zebra finch Taeniopygia guttata provides new insights into avian genome evolution. *Genetics* **179:** 651-667.

Stark, G.R. 1986. DNA amplification in drug resistant cells and in tumours. *Cancer Surv* **5:** 1-23.

Stewart, E.A., K.B. McKusick, A. Aggarwal, E. Bajorek, S. Brady, A. Chu, N. Fang, D. Hadley, M. Harris, S. Hussain, R. Lee, A. Maratukulam, K. O'Connor, S. Perkins, M. Piercy, F. Qin, T. Reif, C. Sanders, X. She, W.L. Sun, P. Tabar, S. Voyticky, S. Cowles, J.B. Fan, C. Mader, J. Quackenbush, R.M. Myers, and D.R. Cox. 1997. An STS-based radiation hybrid map of the human genome. *Genome Res* **7:** 422-433.

Stiglec, R., T. Ezaz, and J.A. Graves. 2007. A new look at the evolution of avian sex chromosomes. *Cytogenet Genome Res* **117:** 103-109.

Storchova, R. and P. Divina. 2006. Nonrandom representation of sex-biased genes on chicken Z chromosome. *J Mol Evol* **63:** 676-681.

Su, F., Y. Osada, M. Ekker, M. Chevrette, A. Shimizu, S. Asakawa, A. Shiohama, T. Sasaki, N. Shimizu, T. Yamanaka, T. Sasado, H. Mitani, R. Geisler, H. Kondoh, and M. Furutani-Seiki. 2007. Radiation hybrid maps of Medaka chromosomes LG 12, 17, and 22. *DNA Res* **14:** 135-140.

Suzuki, S., N. Ono, C. Furusawa, B.W. Ying, and T. Yomo. 2011. Comparison of sequence reads obtained from three next-generation sequencing platforms. *PLoS One* **6:** e19534.

Swayne, D.E. and D.L. Suarez. 2000. Highly pathogenic avian influenza. *Rev Sci Tech* **19:** 463-482.

Tai, C. and R. Rouvier. 1998. Crossbreeding effect on sexual dimorphism of body weight in intergeneric hybrids obtained between Muscovy and Pekin duck. *Genetics Selection Evolution***:** 163-170.

Takahasi, K., M. Yoneyama, T. Nishihori, R. Hirai, H. Kumeta, R. Narita, M. Gale, Jr., F. Inagaki, and T. Fujita. 2008. Nonself RNA-sensing mechanism of RIG-I helicase and activation of antiviral immune responses. *Mol Cell* **29:** 428-440.

Telenius, H., N.P. Carter, C.E. Bebb, M. Nordenskjold, B.A. Ponder, and A. Tunnacliffe. 1992. Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. *Genomics* **13:** 718-725.

Teranishi, M., Y. Shimada, T. Hori, O. Nakabayashi, T. Kikuchi, T. Macleod, R. Pym, B. Sheldon, I. Solovei, H. Macgregor, and S. Mizuno. 2001. Transcripts of the MHM region on the chicken Z chromosome accumulate as non-coding RNA in the nucleus of female cells adjacent to the DMRT1 locus. *Chromosome Res* **9:** 147-165.

Torrents, D., M. Suyama, E. Zdobnov, and P. Bork. 2003. A genome-wide survey of human pseudogenes. *Genome Res* **13:** 2559-2567.

Towers, G.J. 2007. The control of viral infection by tripartite motif proteins and cyclophilin A. *Retrovirology* **4:** 40.

Travers, K.J., C.S. Chin, D.R. Rank, J.S. Eid, and S.W. Turner. 2010. A flexible and efficient template format for circular consensus sequencing and SNP detection. *Nucleic Acids Res* **38:** e159.

Turcatti, G., A. Romieu, M. Fedurco, and A.P. Tairi. 2008. A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis. *Nucleic Acids Res* **36:** e25.

Van Etten, W.J., R.G. Steen, H. Nguyen, A.B. Castle, D.K. Slonim, B. Ge, C. Nusbaum, G.D. Schuler, E.S. Lander, and T.J. Hudson. 1999. Radiation hybrid map of the mouse genome. *Nat Genet* **22:** 384-387.

van Tuinen, M. and G.J. Dyke. 2004. Calibration of galliform molecular clocks using multiple fossils and genetic partitions. *Mol Phylogenet Evol* **30:** 74-86.

van Tuinen, M. and S.B. Hedges. 2001. Calibration of avian molecular clocks. *Mol Biol Evol* **18:** 206-213.

Vandergon, T.L. and M. Reitman. 1994. Evolution of chicken repeat 1 (CR1) elements: evidence for ancient subfamilies and multiple progenitors. *Mol Biol Evol* **11:** 886-898.

Venter, J.C. M.D. Adams E.W. Myers P.W. Li R.J. Mural G.G. Sutton H.O. Smith M. Yandell C.A. Evans R.A. Holt J.D. Gocayne P. Amanatides R.M. Ballew D.H. Huson J.R. Wortman Q. Zhang C.D. Kodira X.H. Zheng L. Chen M. Skupski G. Subramanian P.D. Thomas J. Zhang G.L. Gabor Miklos C. Nelson S. Broder A.G. Clark J. Nadeau V.A. McKusick N. Zinder A.J. Levine R.J. Roberts M. Simon C. Slayman M. Hunkapiller R. Bolanos A. Delcher I. Dew D. Fasulo M. Flanigan L. Florea A. Halpern S. Hannenhalli S. Kravitz S. Levy C. Mobarry K. Reinert K. Remington J. Abu-Threideh E. Beasley K. Biddick V. Bonazzi R. Brandon M. Cargill I. Chandramouliswaran R. Charlab K. Chaturvedi Z. Deng V. Di Francesco P. Dunn K. Eilbeck C. Evangelista A.E. Gabrielian W. Gan W. Ge F. Gong Z. Gu P. Guan T.J. Heiman M.E. Higgins R.R. Ji Z. Ke K.A. Ketchum Z. Lai Y. Lei Z. Li J. Li Y. Liang X. Lin F. Lu G.V. Merkulov N. Milshina H.M. Moore A.K. Naik V.A. Narayan B. Neelam D. Nusskern D.B. Rusch S. Salzberg W. Shao B. Shue J. Sun Z. Wang A.

Wang X. Wang J. Wang M. Wei R. Wides C. Xiao C. Yan A. Yao J. Ye M. Zhan W. Zhang H. Zhang Q. Zhao L. Zheng F. Zhong W. Zhong S. Zhu S. Zhao D. Gilbert S. Baumhueter G. Spier C. Carter A. Cravchik T. Woodage F. Ali H. An A. Awe D. Baldwin H. Baden M. Barnstead I. Barrow K. Beeson D. Busam A. Carver A. Center M.L. Cheng L. Curry S. Danaher L. Davenport R. Desilets S. Dietz K. Dodson L. Doup S. Ferriera N. Garg A. Gluecksmann B. Hart J. Haynes C. Haynes C. Heiner S. Hladun D. Hostin J. Houck T. Howland C. Ibegwam J. Johnson F. Kalush L. Kline S. Koduru A. Love F. Mann D. May S. McCawley T. McIntosh I. McMullen M. Moy L. Moy B. Murphy K. Nelson C. Pfannkoch E. Pratts V. Puri H. Qureshi M. Reardon R. Rodriguez Y.H. Rogers D. Romblad B. Ruhfel R. Scott C. Sitter M. Smallwood E. Stewart R. Strong E. Suh R. Thomas N.N. Tint S. Tse C. Vech G. Wang J. Wetter S. Williams M. Williams S. Windsor E. Winn-Deen K. Wolfe J. Zaveri K. Zaveri J.F. Abril R. Guigo M.J. Campbell K.V. Sjolander B. Karlak A. Kejariwal H. Mi B. Lazareva T. Hatton A. Narechania K. Diemer A. Muruganujan N. Guo S. Sato V. Bafna S. Istrail R. Lippert R. Schwartz B. Walenz S. Yooseph D. Allen A. Basu J. Baxendale L. Blick M. Caminha J. Carnes-Stine P. Caulk Y.H. Chiang M. Coyne C. Dahlke A. Mays M. Dombroski M. Donnelly D. Ely S. Esparham C. Fosler H. Gire S. Glanowski K. Glasser A. Glodek M. Gorokhov K. Graham B. Gropman M. Harris J. Heil S. Henderson J. Hoover D. Jennings C. Jordan J. Jordan J. Kasha L. Kagan C. Kraft A. Levitsky M. Lewis X. Liu J. Lopez D. Ma W. Majoros J. McDaniel S. Murphy M. Newman T. Nguyen N. Nguyen M. Nodell S. Pan J. Peck M. Peterson W. Rowe R. Sanders J. Scott M. Simpson T. Smith A. Sprague T. Stockwell R. Turner E. Venter M. Wang M. Wen D. Wu M. Wu A. Xia A. Zandieh and X. Zhu. 2001. The sequence of the human genome. *Science* **291:** 1304-1351.

Vignal, A., D. Milan, M. SanCristobal, and A. Eggen. 2002. A review on SNP and other types of molecular markers and their use in animal genetics. *Genet Sel Evol* **34:** 275-305.

Vignaux, F., C. Hitte, C. Priat, J.C. Chuat, C. Andre, and F. Galibert. 1999. Construction and optimization of a dog whole-genome radiation hybrid panel. *Mamm Genome* **10:** 888-894.

Volker, M., N. Backstrom, B.M. Skinner, E.J. Langley, S.K. Bunzey, H. Ellegren, and D.K. Griffin. 2010. Copy number variation, chromosome rearrangement, and their association with recombination during avian evolution. *Genome Res* **20:** 503-511.

Wallis, J.W., J. Aerts, M.A. Groenen, R.P. Crooijmans, D. Layman, T.A. Graves, D.E. Scheer, C. Kremitzki, M.J. Fedele, N.K. Mudd, M. Cardenas, J. Higginbotham, J. Carter, R. McGrane, T. Gaige, K. Mead, J. Walker, D. Albracht, J. Davito, S.P. Yang, S. Leong, A. Chinwalla, M. Sekhon, K. Wylie, J. Dodgson, M.N. Romanov, H. Cheng, P.J. de Jong, K. Osoegawa, M. Nefedov, H. Zhang, J.D. McPherson, M. Krzywinski, J. Schein, L. Hillier, E.R. Mardis, R.K. Wilson, and W.C. Warren. 2004. A physical map of the chicken genome. *Nature* **432:** 761-764.

Walter, M.A. and P.N. Goodfellow. 1993. Radiation hybrids: irradiation and fusion gene transfer. *Trends Genet* **9:** 352-356.

Walter, M.A., D.J. Spillett, P. Thomas, J. Weissenbach, and P.N. Goodfellow. 1994. A method for constructing radiation hybrid maps of whole genomes. *Nat Genet* **7:** 22-28.

Warren, W.C., D.F. Clayton, H. Ellegren, A.P. Arnold, L.W. Hillier, A. Kunstner, S. Searle, S. White, A.J. Vilella, S. Fairley, A. Heger, L. Kong, C.P. Ponting, E.D. Jarvis, C.V. Mello, P. Minx, P. Lovell, T.A. Velho, M. Ferris, C.N. Balakrishnan, S. Sinha, C. Blatti, S.E. London, Y. Li, Y.C. Lin, J. George, J. Sweedler, B. Southey, P. Gunaratne, M. Watson, K. Nam, N. Backstrom, L. Smeds, B. Nabholz, Y. Itoh, O. Whitney, A.R. Pfenning, J. Howard, M. Volker, B.M. Skinner, D.K. Griffin, L. Ye, W.M. McLaren, P. Flicek, V. Quesada, G. Velasco, C. Lopez-Otin, X.S. Puente, T. Olender, D. Lancet, A.F. Smit, R. Hubley, M.K. Konkel, J.A. Walker, M.A. Batzer, W. Gu, D.D. Pollock, L. Chen, Z. Cheng, E.E. Eichler, J. Stapley, J. Slate, R. Ekblom, T. Birkhead, T. Burke, D. Burt, C. Scharff, I. Adam, H. Richard, M. Sultan, A. Soldatov, H. Lehrach, S.V. Edwards, S.P. Yang, X. Li, T. Graves, L. Fulton, J. Nelson, A. Chinwalla, S. Hou, E.R. Mardis, and R.K. Wilson. 2010. The genome of a songbird. *Nature* **464:** 757-762.

Watanabe, T.K., M.T. Bihoreau, L.C. McCarthy, S.L. Kiguwa, H. Hishigaki, A. Tsuji, J. Browne, Y. Yamasaki, A. Mizoguchi-Miyakita, K. Oga, T. Ono, S. Okuno, N. Kanemoto, E. Takahashi, K. Tomita, H. Hayashi, M. Adachi, C. Webber, M. Davis, S. Kiel, C. Knights, A. Smith, R. Critcher, J. Miller, T. Thangarajah, P.J. Day, J.R. Hudson, Jr., Y. Irie, T. Takagi, Y. Nakamura, P.N. Goodfellow, G.M. Lathrop, A. Tanigami, and M.R. James. 1999. A radiation hybrid map of the rat genome containing 5,255 markers. *Nat Genet* **22:** 27-36.

Waterston, R.H. K. Lindblad-Toh E. Birney J. Rogers J.F. Abril P. Agarwal R. Agarwala R. Ainscough M. Alexandersson P. An S.E. Antonarakis J. Attwood R. Baertsch J. Bailey K. Barlow S. Beck E. Berry B. Birren T. Bloom P. Bork M. Botcherby N. Bray M.R. Brent D.G. Brown S.D. Brown C. Bult J. Burton J. Butler R.D. Campbell P. Carninci S. Cawley F. Chiaromonte A.T. Chinwalla D.M. Church M. Clamp C. Clee F.S. Collins L.L. Cook R.R. Copley A. Coulson O. Couronne J. Cuff V. Curwen T. Cutts M. Daly R. David J. Davies K.D. Delehaunty J. Deri E.T. Dermitzakis C. Dewey N.J. Dickens M. Diekhans S. Dodge I. Dubchak D.M. Dunn S.R. Eddy L. Elnitski R.D. Emes P. Eswara E. Eyras A. Felsenfeld G.A. Fewell P. Flicek K. Foley W.N. Frankel L.A. Fulton R.S. Fulton T.S. Furey D. Gage R.A. Gibbs G. Glusman S. Gnerre N. Goldman L. Goodstadt D. Grafham T.A. Graves E.D. Green S. Gregory R. Guigo M. Guyer R.C. Hardison D. Haussler Y. Hayashizaki L.W. Hillier A. Hinrichs W. Hlavina T. Holzer F. Hsu A. Hua T. Hubbard A. Hunt I. Jackson D.B. Jaffe L.S. Johnson M. Jones T.A. Jones A. Joy M. Kamal E.K. Karlsson D. Karolchik A. Kasprzyk J. Kawai E. Keibler C. Kells W.J. Kent A. Kirby D.L. Kolbe I. Korf R.S. Kucherlapati E.J. Kulbokas D. Kulp T. Landers J.P. Leger S. Leonard I. Letunic R.

Levine J. Li M. Li C. Lloyd S. Lucas B. Ma D.R. Maglott E.R. Mardis L. Matthews E. Mauceli J.H. Mayer M. McCarthy W.R. McCombie S. McLaren K. McLay J.D. McPherson J. Meldrim B. Meredith J.P. Mesirov W. Miller T.L. Miner E. Mongin K.T. Montgomery M. Morgan R. Mott J.C. Mullikin D.M. Muzny W.E. Nash J.O. Nelson M.N. Nhan R. Nicol Z. Ning C. Nusbaum M.J. O'Connor Y. Okazaki K. Oliver E. Overton-Larty L. Pachter G. Parra K.H. Pepin J. Peterson P. Pevzner R. Plumb C.S. Pohl A. Poliakov T.C. Ponce C.P. Ponting S. Potter M. Quail A. Reymond B.A. Roe K.M. Roskin E.M. Rubin A.G. Rust R. Santos V. Sapojnikov B. Schultz J. Schultz M.S. Schwartz S. Schwartz C. Scott S. Seaman S. Searle T. Sharpe A. Sheridan R. Shownkeen S. Sims J.B. Singer G. Slater A. Smit D.R. Smith B. Spencer A. Stabenau N. Stange-Thomann C. Sugnet M. Suyama G. Tesler J. Thompson D. Torrents E. Trevaskis J. Tromp C. Ucla A. Ureta-Vidal J.P. Vinson A.C. Von Niederhausern C.M. Wade M. Wall R.J. Weber R.B. Weiss M.C. Wendl A.P. West K. Wetterstrand R. Wheeler S. Whelan J. Wierzbowski D. Willey S. Williams R.K. Wilson E. Winter K.C. Worley D. Wyman S. Yang S.P. Yang E.M. Zdobnov M.C. Zody and E.S. Lander. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420:** 520-562.

Westerveld, A., R.P. Visser, P. Meera Khan, and D. Bootsma. 1971. Loss of human genetic markers in man--Chinese hamster somatic cell hybrids. *Nat New Biol* **234:** 20-24.

Wicker, T., J.S. Robertson, S.R. Schulze, F.A. Feltus, V. Magrini, J.A. Morrison, E.R. Mardis, R.K. Wilson, D.G. Peterson, A.H. Paterson, and R. Ivarie. 2005. The repetitive landscape of the chicken genome. *Genome Res* **15:** 126-136.

Wicker, T., N. Stein, L. Albar, C. Feuillet, E. Schlagenhauf, and B. Keller. 2001. Analysis of a contiguous 211 kb sequence in diploid wheat (Triticum monococcum L.) reveals multiple mechanisms of genome evolution. *Plant J* **26:** 307-316.

Wiedmann, R.T., D.J. Nonneman, and J.W. Keele. 2006. Novel porcine repetitive elements. *BMC Genomics* **7:** 304.

Williams, J.L., A. Eggen, L. Ferretti, C.J. Farr, M. Gautier, G. Amati, G. Ball, T. Caramorr, R. Critcher, S. Costa, P. Hextall, D. Hills, A. Jeulin, S.L. Kiguwa, O. Ross, A.L. Smith, K. Saunier, B. Urquhart, and D. Waddington. 2002. A bovine whole-genome radiation hybrid panel and outline map. *Mamm Genome* **13:** 469-474.

Womack, J.E., J.S. Johnson, E.K. Owens, C.E. Rexroad, 3rd, J. Schlapfer, and Y.P. Yang. 1997. A whole-genome radiation hybrid panel for bovine gene mapping. *Mamm Genome* **8:** 854-856.

Wong, G.K. B. Liu J. Wang Y. Zhang X. Yang Z. Zhang Q. Meng J. Zhou D. Li J. Zhang P. Ni S. Li L. Ran H. Li R. Li H. Zheng W. Lin G. Li X. Wang W. Zhao J. Li C. Ye M. Dai J. Ruan Y. Zhou Y. Li X. He X. Huang W. Tong J. Chen J. Ye C. Chen N. Wei L. Dong F. Lan Y. Sun Z. Yang Y. Yu Y. Huang D. He Y. Xi D. Wei Q. Qi W. Li J. Shi M. Wang F. Xie X. Zhang P. Wang Y. Zhao N. Li N. Yang W. Dong S. Hu C. Zeng

W. Zheng B. Hao L.W. Hillier S.P. Yang W.C. Warren R.K. Wilson M. Brandstrom H. Ellegren R.P. Crooijmans J.J. van der Poel H. Bovenhuis M.A. Groenen I. Ovcharenko L. Gordon L. Stubbs S. Lucas T. Glavina A. Aerts P. Kaiser L. Rothwell J.R. Young S. Rogers B.A. Walker A. van Hateren J. Kaufman N. Bumstead S.J. Lamont H. Zhou P.M. Hocking D. Morrice D.J. de Koning A. Law N. Bartley D.W. Burt H. Hunt H.H. Cheng U. Gunnarsson P. Wahlberg L. Andersson E. Kindlund M.T. Tammi B. Andersson C. Webber C.P. Ponting I.M. Overton P.E. Boardman H. Tang S.J. Hubbard S.A. Wilson J. Yu and H. Yang. 2004. A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. *Nature* **432:** 717-722.

Wu, C.H., K. Nomura, T. Goldammer, T. Hadfield, J.E. Womack, and N.E. Cockett. 2007. An ovine whole-genome radiation hybrid panel used to construct an RH map of ovine chromosome 9. *Anim Genet* **38:** 534-536.

Wu, R. and E. Taylor. 1971. Nucleotide sequence analysis of DNA. II. Complete nucleotide sequence of the cohesive ends of bacteriophage lambda DNA. *J Mol Biol* **57:** 491-511.

Wucheng, B. 1988. The research on the origin of the house-duck in China. In *The Satellite Conference for the18th World's Poultry Congress*, pp. 125-129. Pergamon Press, Oxford, Beijing, China.

Xia, Q., Y. Guo, Z. Zhang, D. Li, Z. Xuan, Z. Li, F. Dai, Y. Li, D. Cheng, R. Li, T. Cheng, T. Jiang, C. Becquet, X. Xu, C. Liu, X. Zha, W. Fan, Y. Lin, Y. Shen, L. Jiang, J. Jensen, I. Hellmann, S. Tang, P. Zhao, H. Xu, C. Yu, G. Zhang, J. Li, J. Cao, S. Liu, N. He, Y. Zhou, H. Liu, J. Zhao, C. Ye, Z. Du, G. Pan, A. Zhao, H. Shao, W. Zeng, P. Wu, C. Li, M. Pan, X. Yin, J. Wang, H. Zheng, W. Wang, X. Zhang, S. Li, H. Yang, C. Lu, R. Nielsen, Z. Zhou, and Z. Xiang. 2009. Complete resequencing of 40 genomes reveals domestication events and genes in silkworm (Bombyx). *Science* **326:** 433-436.

Xu, X., Y. Hou, X. Yin, L. Bao, A. Tang, L. Song, F. Li, S. Tsang, K. Wu, H. Wu, W. He, L. Zeng, M. Xing, R. Wu, H. Jiang, X. Liu, D. Cao, G. Guo, X. Hu, Y. Gui, Z. Li, W. Xie, X. Sun, M. Shi, Z. Cai, B. Wang, M. Zhong, J. Li, Z. Lu, N. Gu, X. Zhang, L. Goodman, L. Bolund, J. Wang, H. Yang, K. Kristiansen, M. Dean, and Y. Li. 2012. Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. *Cell* **148:** 886-895.

Ye, L., L.W. Hillier, P. Minx, N. Thane, D.P. Locke, J.C. Martin, L. Chen, M. Mitreva, J.R. Miller, K.V. Haub, D.J. Dooling, E.R. Mardis, R.K. Wilson, G.M. Weinstock, and W.C. Warren. 2011. A vertebrate case study of the quality of assemblies derived from next-generation sequences. *Genome Biol* **12:** R31.

Yerle, M., P. Pinton, A. Robic, A. Alfonso, Y. Palvadeau, C. Delcros, R. Hawken, L. Alexander, C. Beattie, L. Schook, D. Milan, and J. Gellin. 1998. Construction of a whole-genome radiation hybrid panel for high-resolution gene mapping in pigs. *Cytogenet Cell Genet* **82:** 182-188.

Yoneyama, M., M. Kikuchi, T. Natsukawa, N. Shinobu, T. Imaizumi, M. Miyagishi, K. Taira, S. Akira, and T. Fujita. 2004. The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. *Nat Immunol* **5:** 730-737.

Yuan, X., M. Zhang, W. Ruan, C. Song, L. Ren, Y. Guo, X. Hu, and N. Li. 2006. Construction and characterization of a duck bacterial artificial chromosome library. *Anim Genet* **37:** 599-600.

Zhang, L., X. Cui, K. Schmitt, R. Hubert, W. Navidi, and N. Arnheim. 1992. Whole genome amplification from a single cell: implications for genetic analysis. *Proc Natl Acad Sci U S A* **89:** 5847-5851.

Zhang, Y., X. Zhang, T.H. O'Hare, W.S. Payne, J.J. Dong, C.F. Scheuring, M. Zhang, J.J. Huang, M.K. Lee, M.E. Delany, H.B. Zhang, and J.B. Dodgson. 2012. A comparative physical map reveals the pattern of chromosomal evolution between the turkey (Meleagris gallopavo) and chicken (Gallus gallus) genomes. *BMC Genomics* **12:** 447.

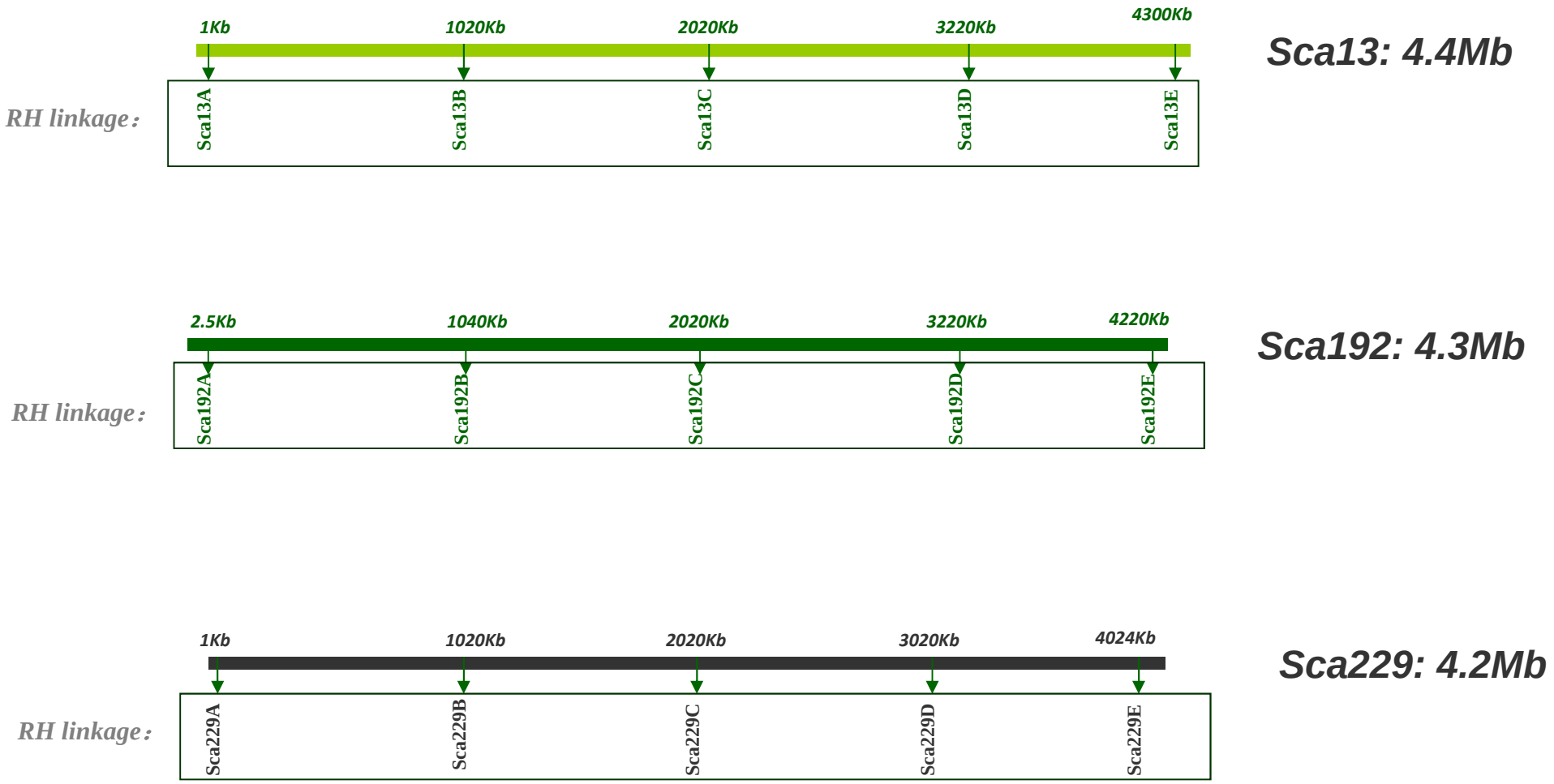
Zhao, S., J. Shetty, L. Hou, A. Delcher, B. Zhu, K. Osoegawa, P. de Jong, W.C. Nierman, R.L. Strausberg, and C.M. Fraser. 2004. Human, mouse, and rat genome large-scale rearrangements: stability versus speciation. *Genome Res* **14:** 1851-1860.

Annex A

Supplementary data to Article 1

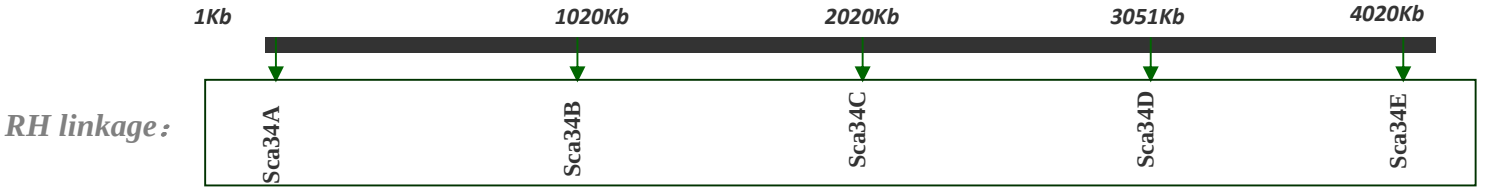
(Chapter III)

13 Largest scaffolds > 4 Mb

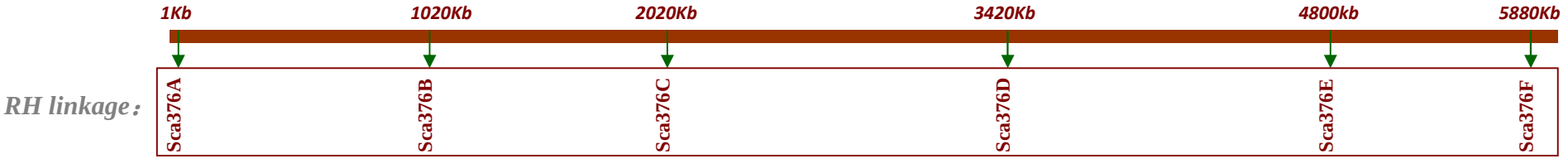


Markers scaXXX… contained in the same box are linked together by RH mapping.

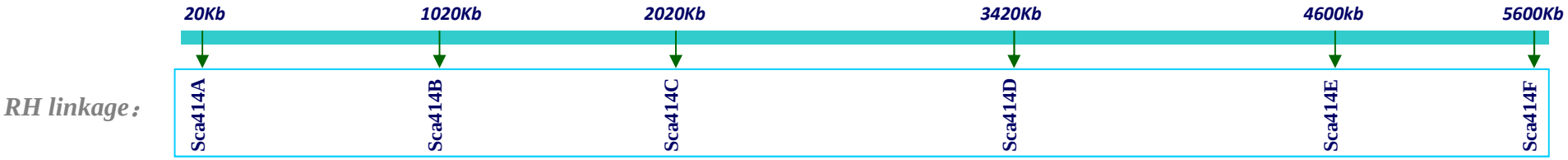
13 Largest scaffolds > 4 Mb



Sca34: 4.1Mb



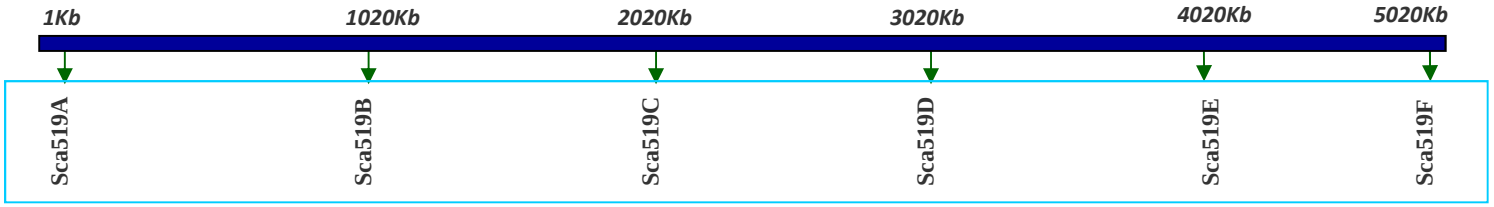
Sca376: 5.9Mb



Sca414: 5.7Mb

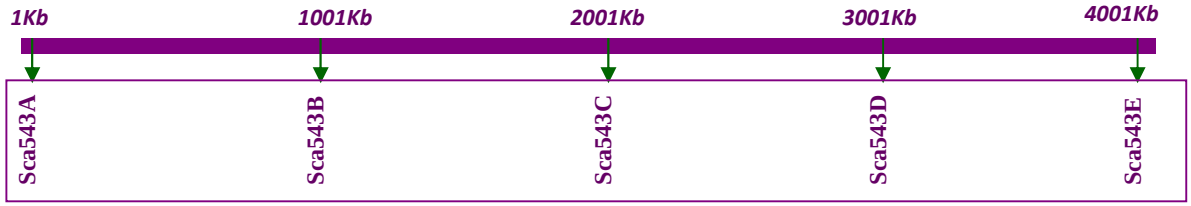
13 Largest scaffolds > 4 Mb

RH linkage：



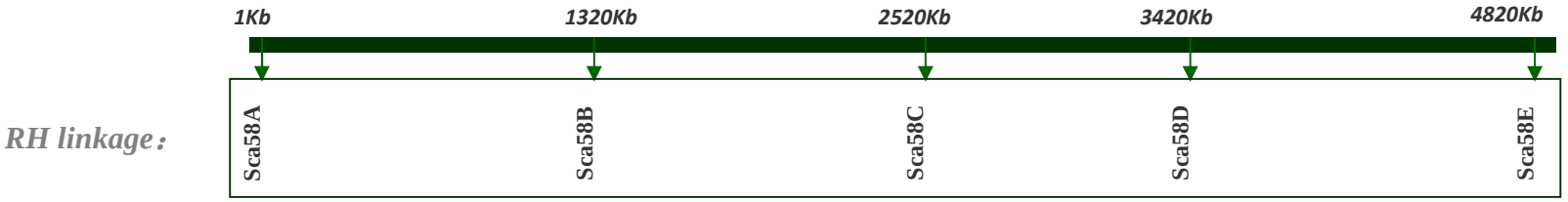
Sca519: 5Mb

RH linkage：

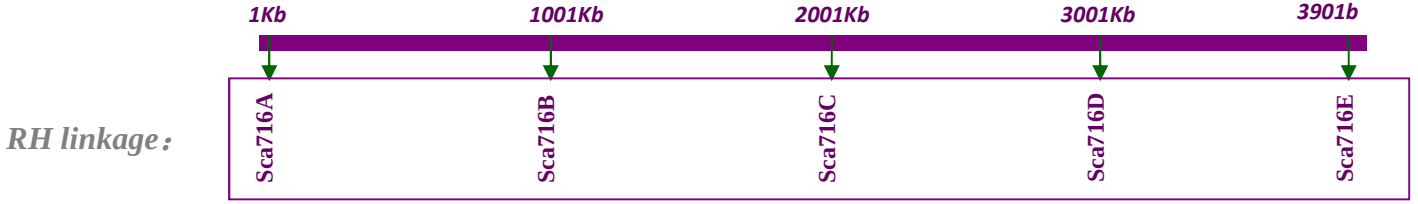


Sca543: 4Mb

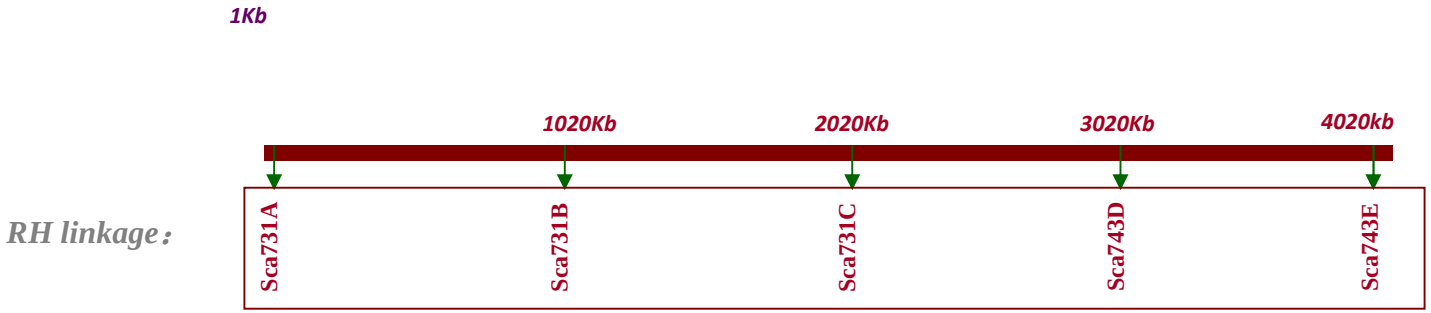
13 Largest scaffolds > 4 Mb



Sca58: 4.9Mb



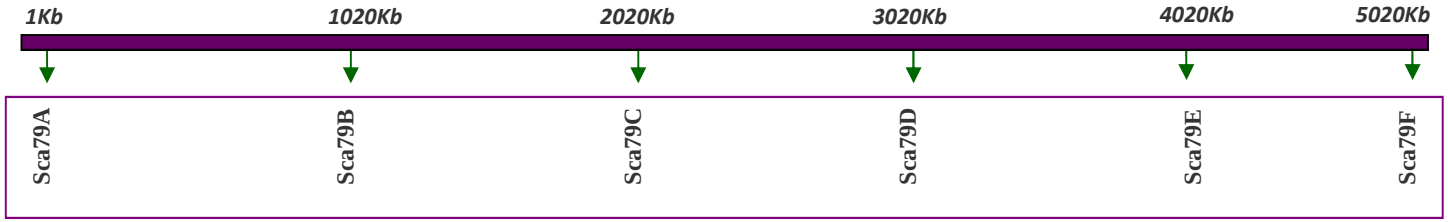
Sca716: 4Mb



Sca731: 4.1Mb

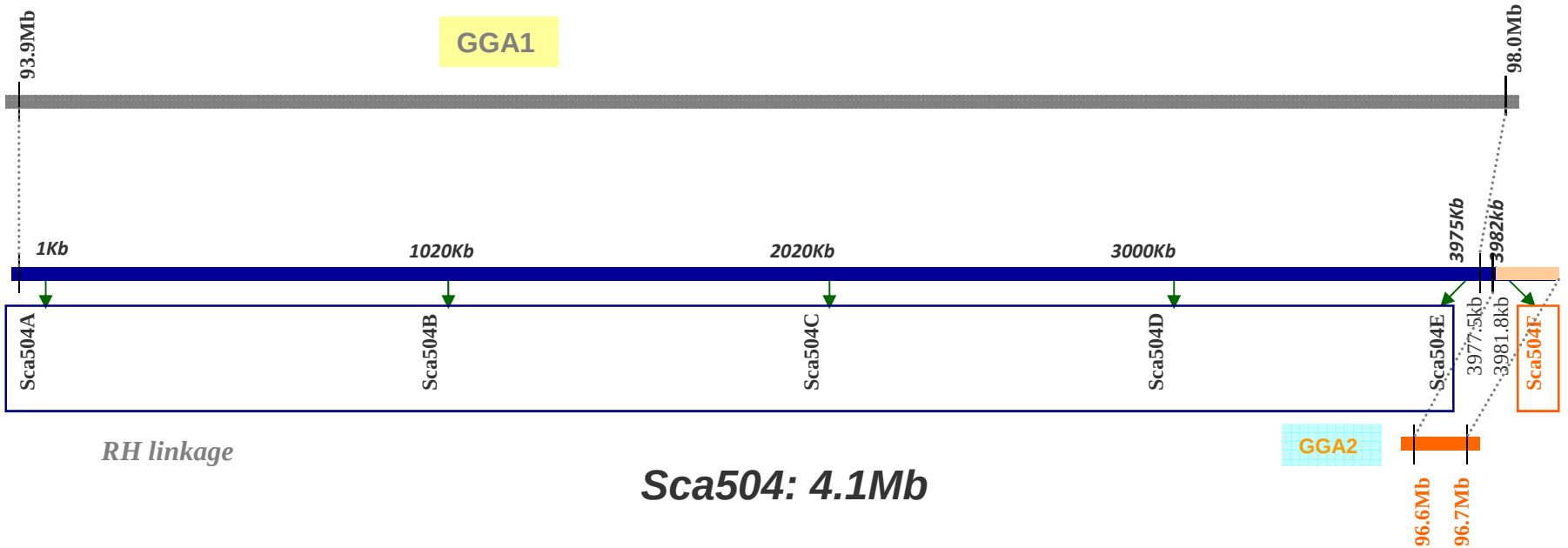
13 Largest scaffolds > 4 Mb

RH linkage：



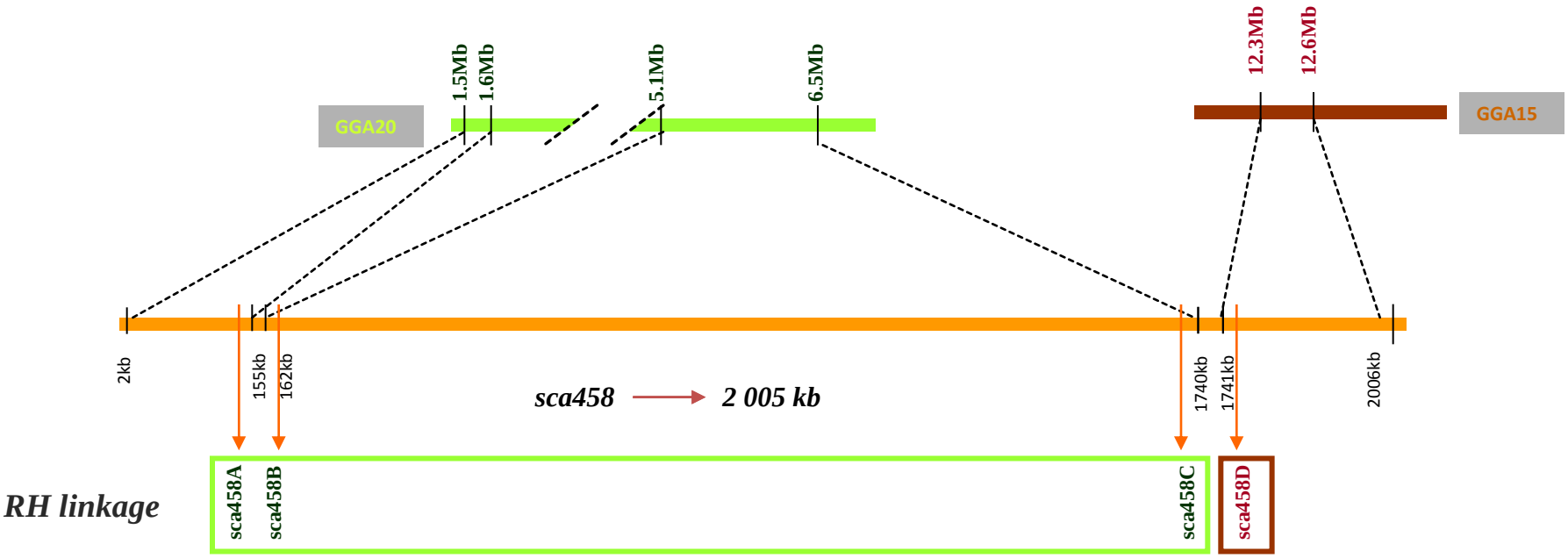
Sca79: 5.1Mb

13 Largest scaffolds > 4 Mb



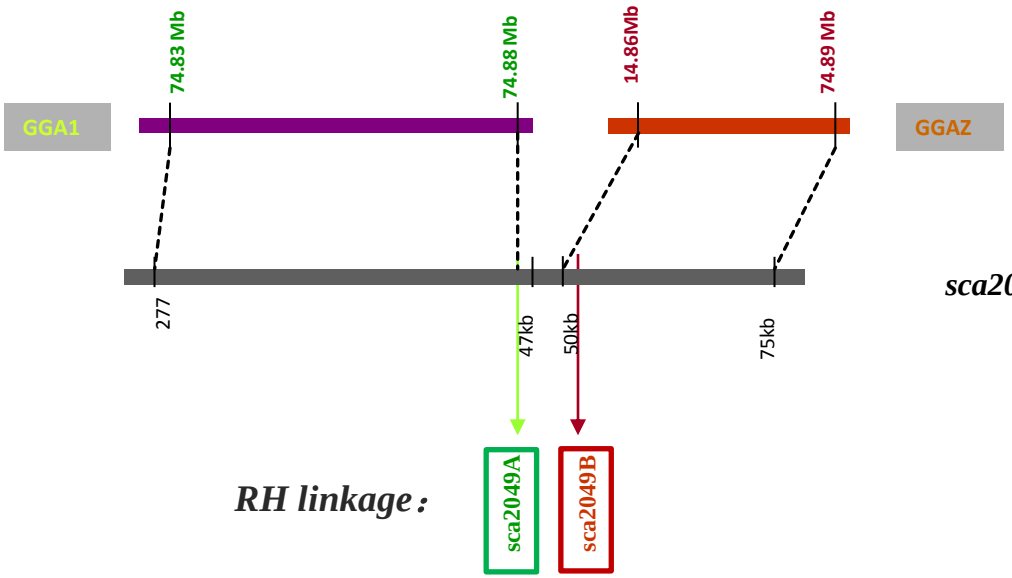
Markers sca504A, sca504B, sca504C, sca504D and sca504E are linked together by RH mapping, whereas marker sca504F is not linked, suggesting the end of the scaffold is misassembled.

Duck scaffolds aligning to two chicken chromosomes



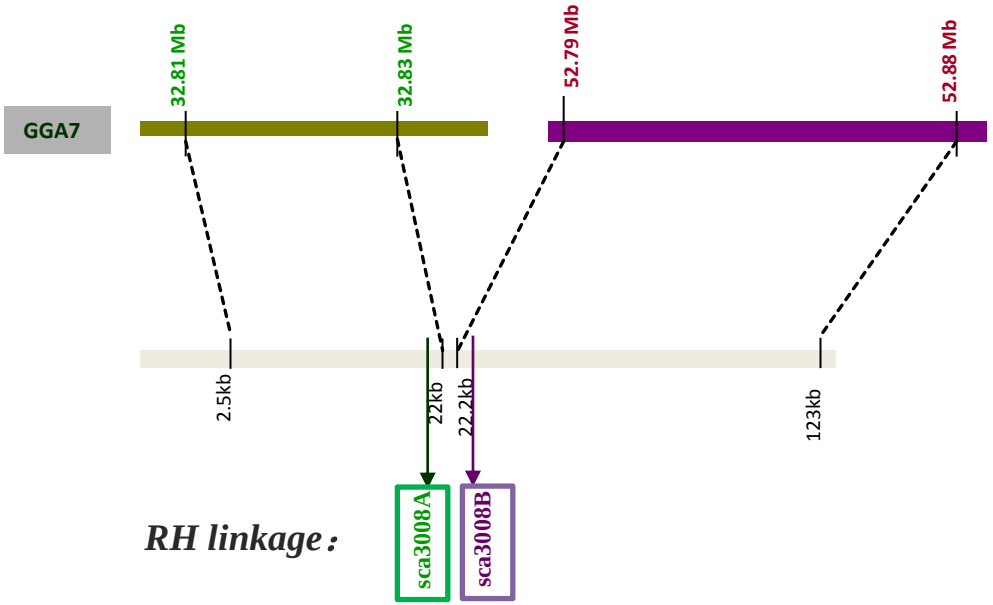
Markers scaXXX… contained in the same box are linked together by RH mapping.

Duck scaffolds aligning to two chicken chromosomes



RH linkage：

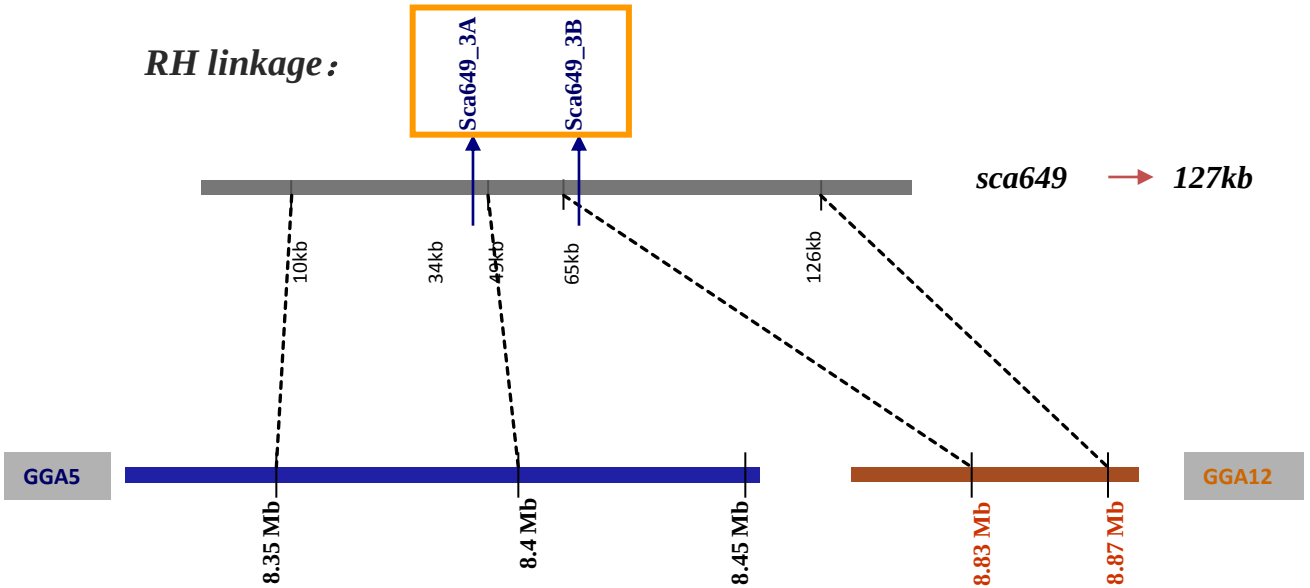
sca2049 → *96kb*



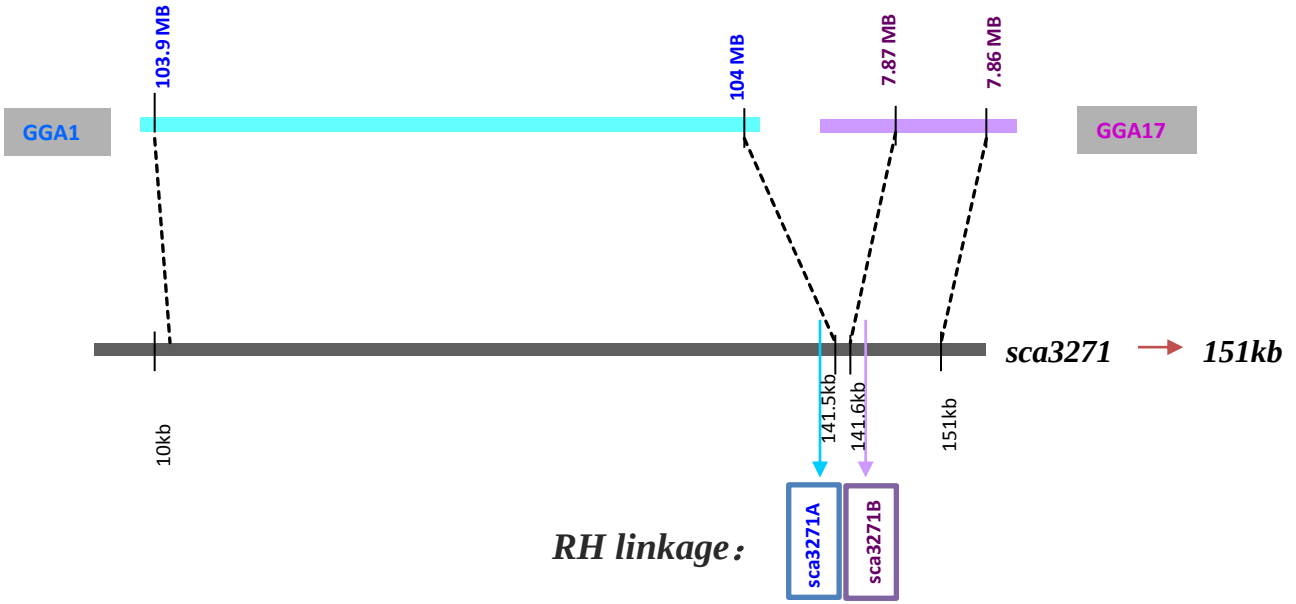
RH linkage：

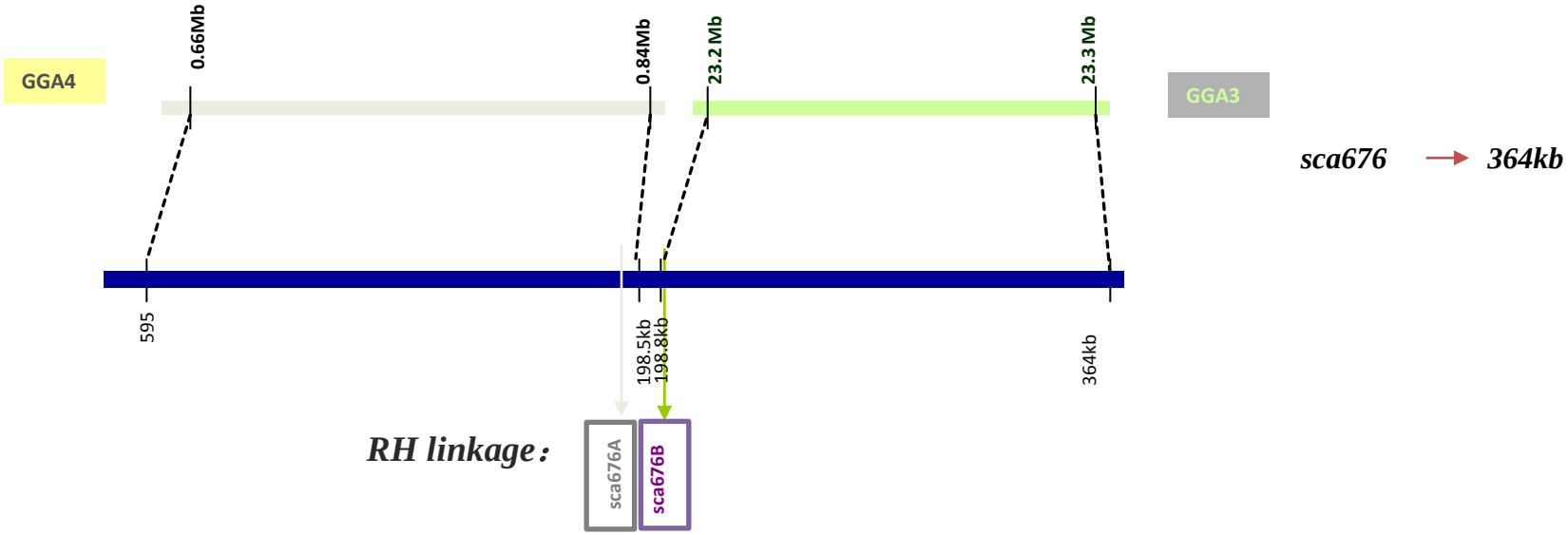
sca3008 → *123kb*

Duck scaffolds aligning to two chicken chromosomes

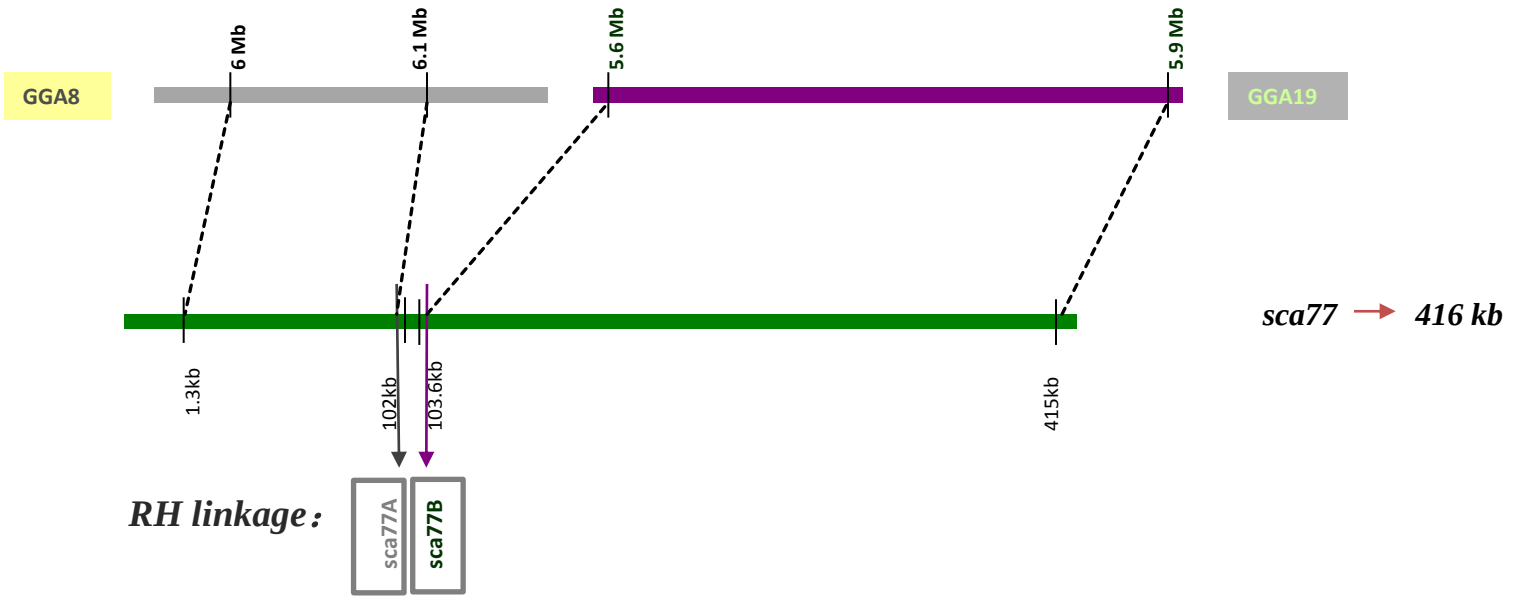


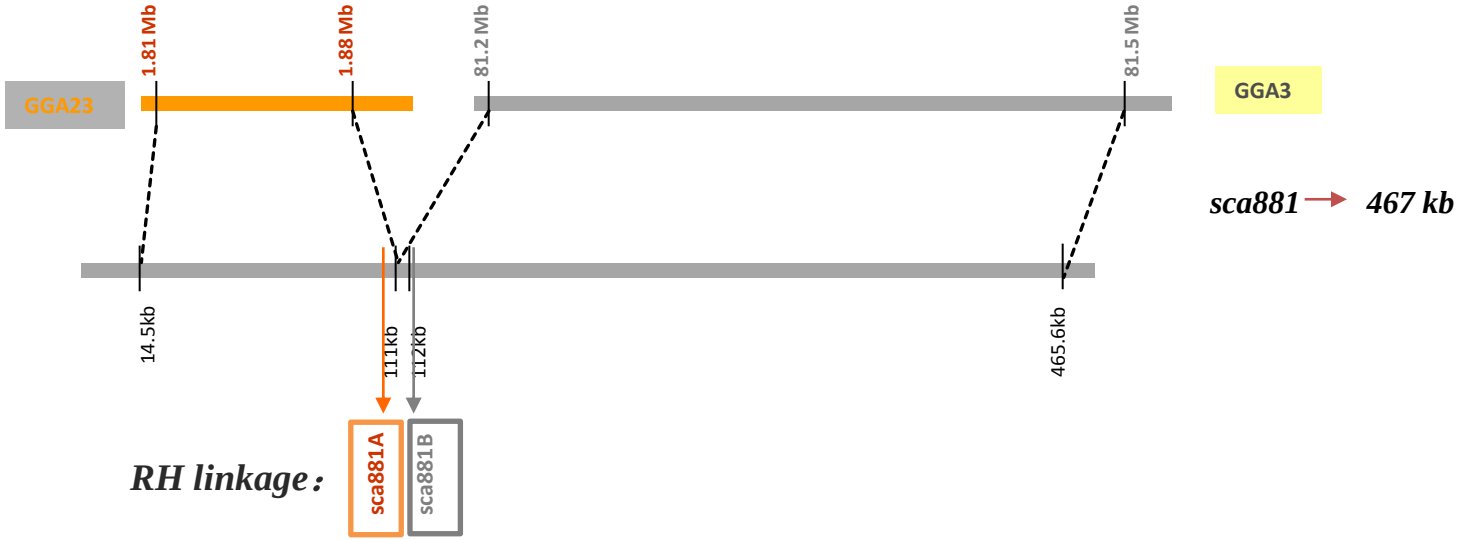
Duck scaffolds aligning to two chicken chromosomes



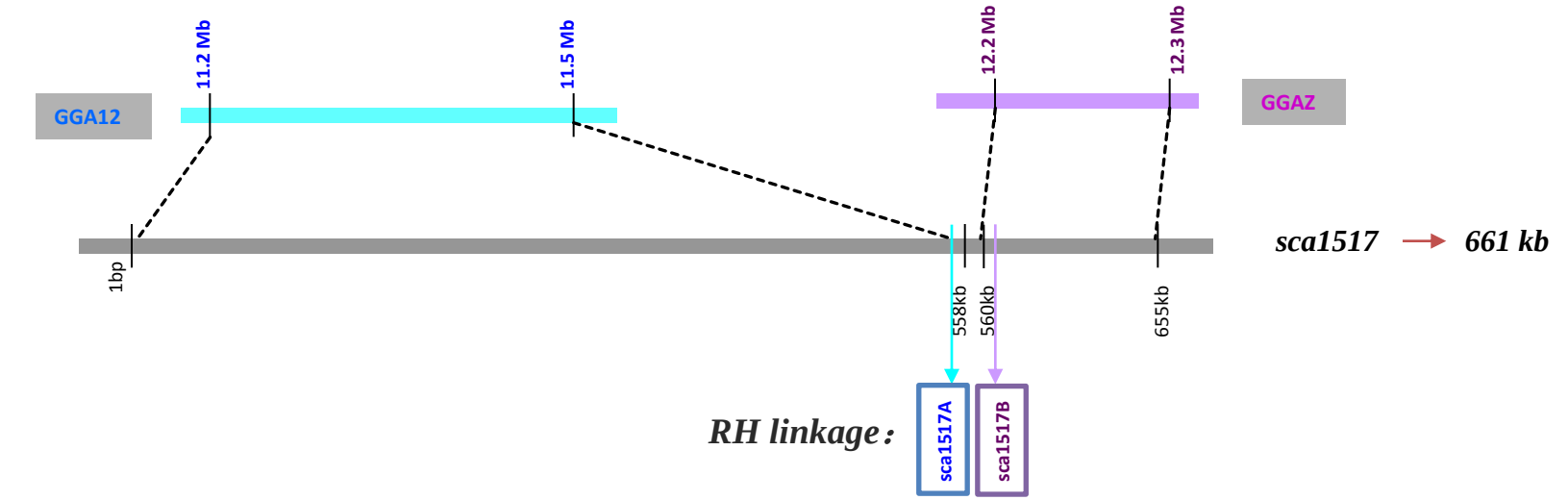


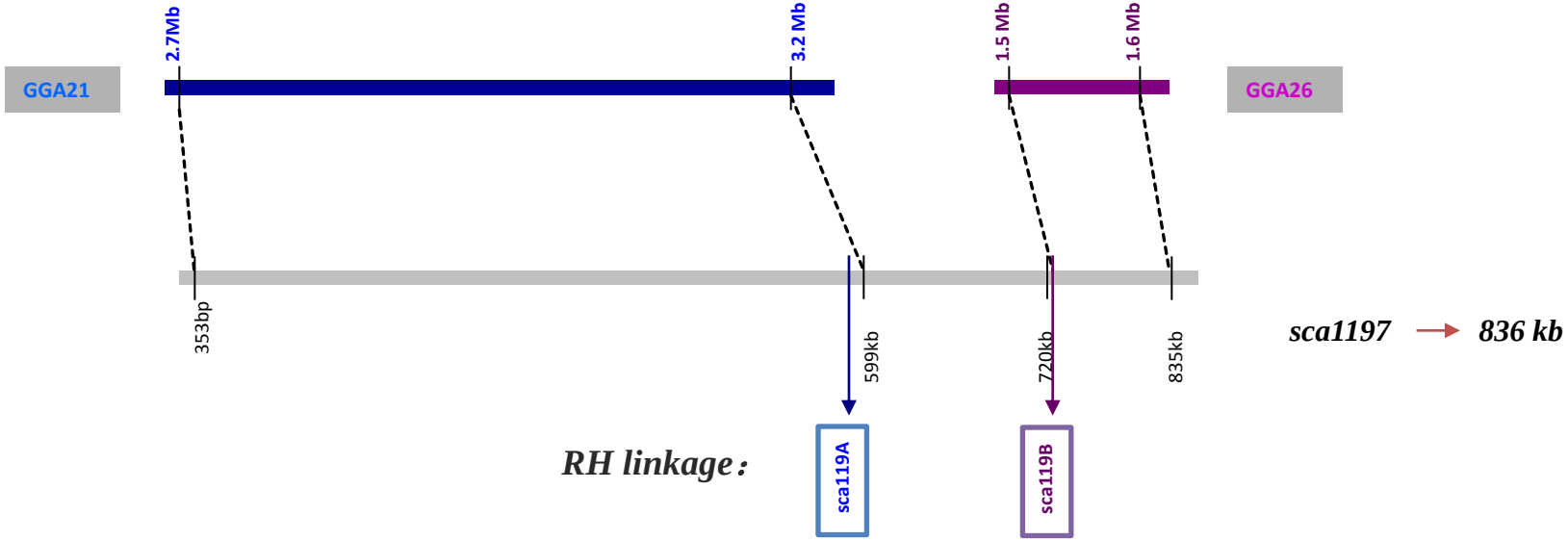
Duck scaffolds aligning to two chicken chromosomes



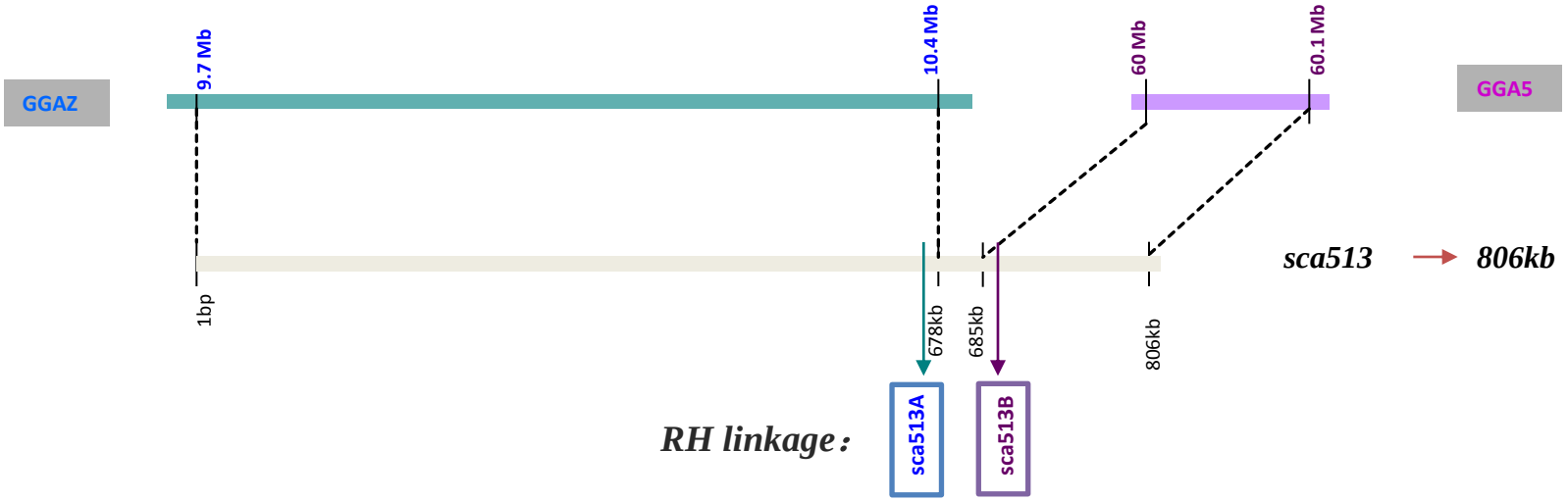


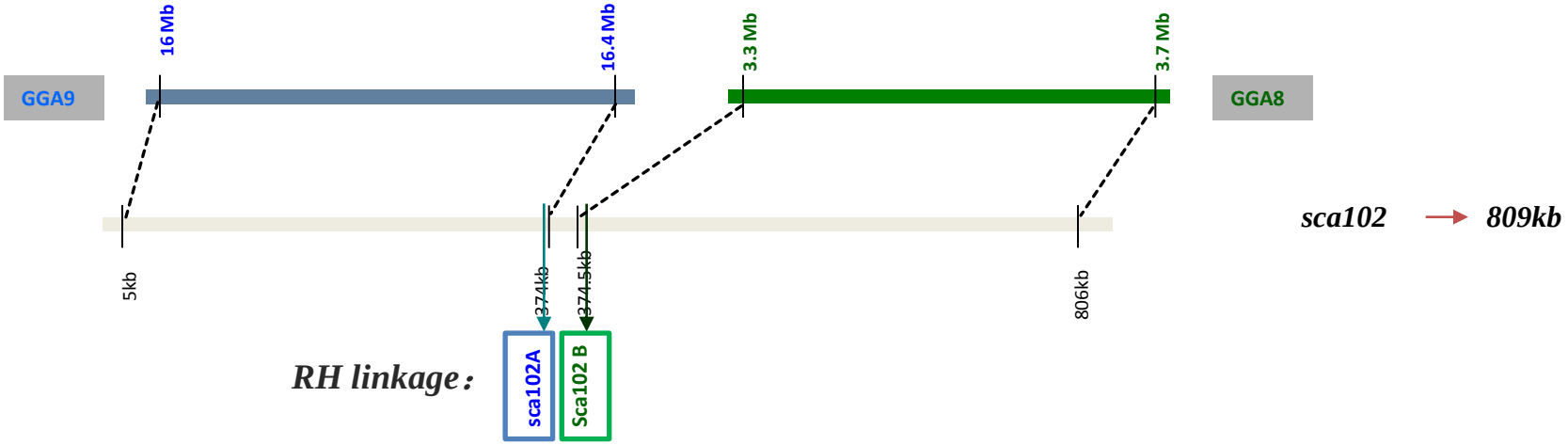
Duck scaffolds aligning to two chicken chromosomes



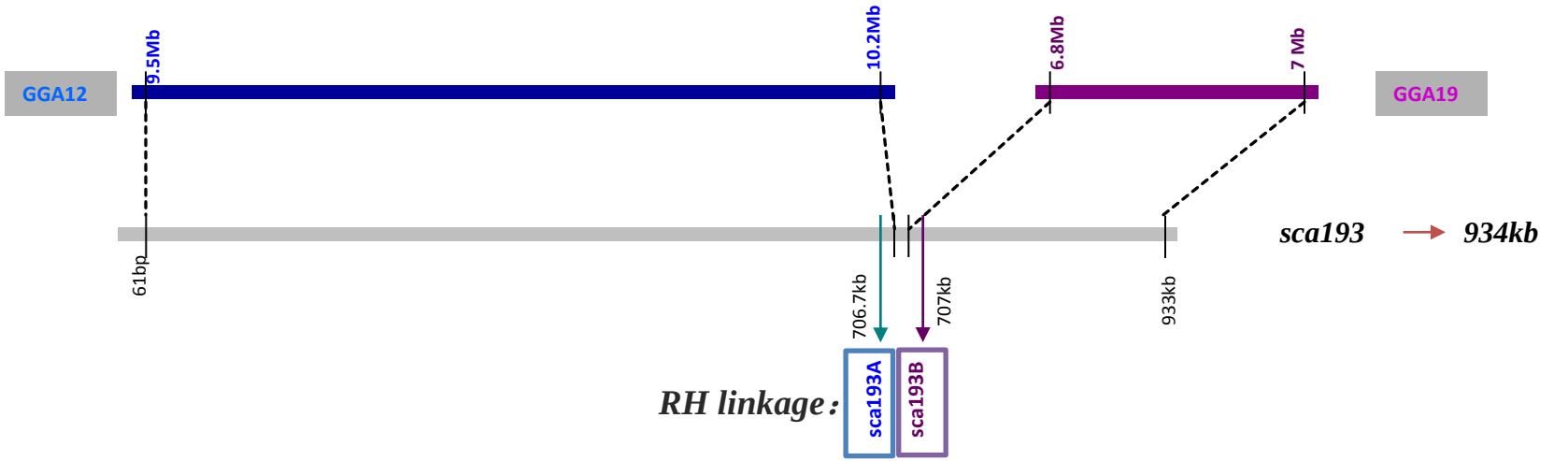


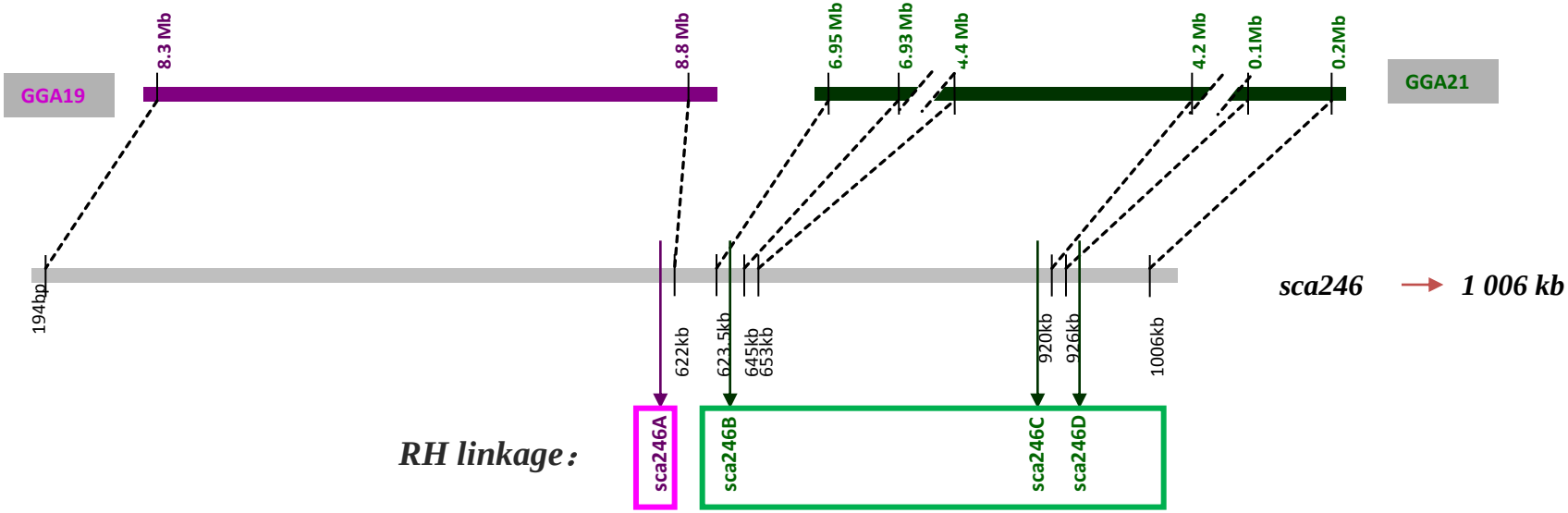
Duck scaffolds aligning to two chicken chromosomes



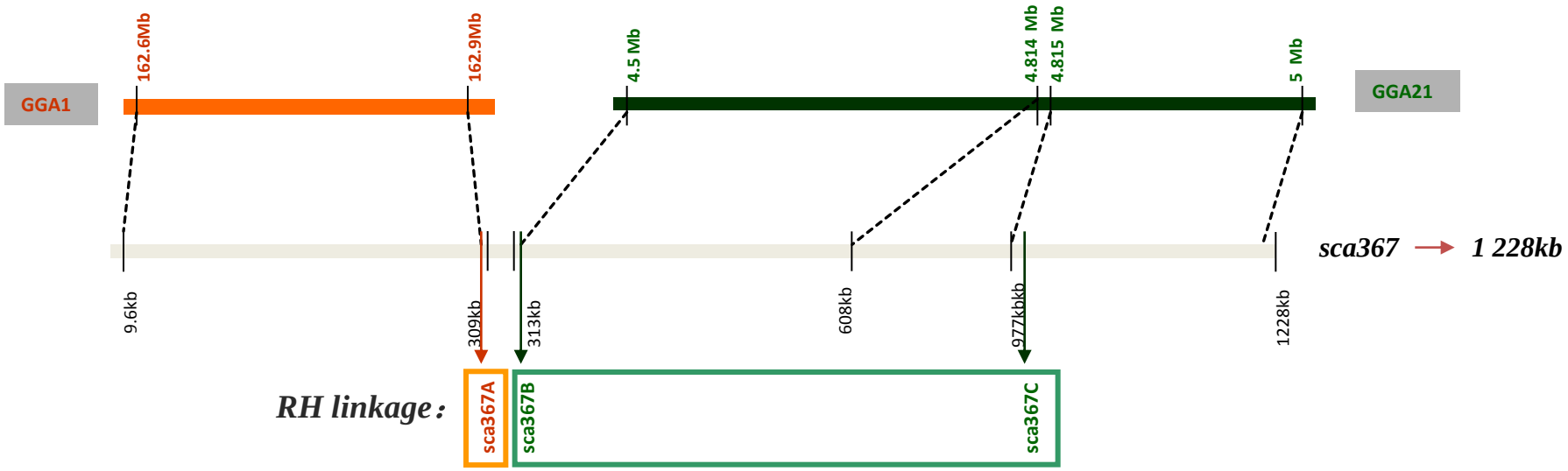


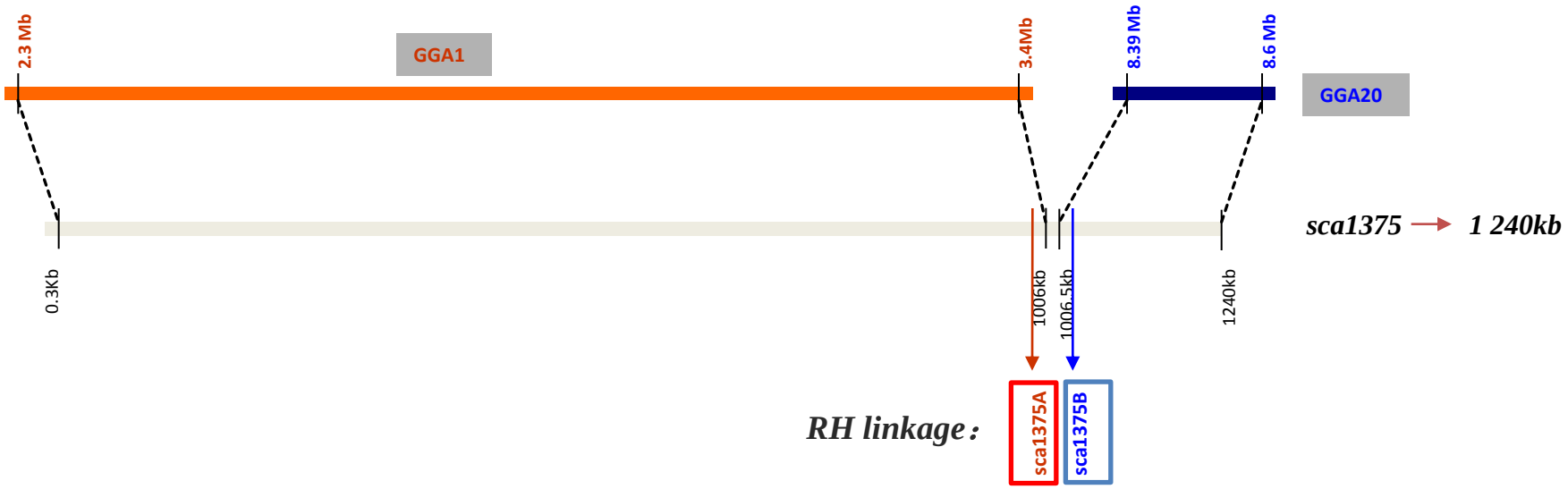
Duck scaffolds aligning to two chicken chromosomes



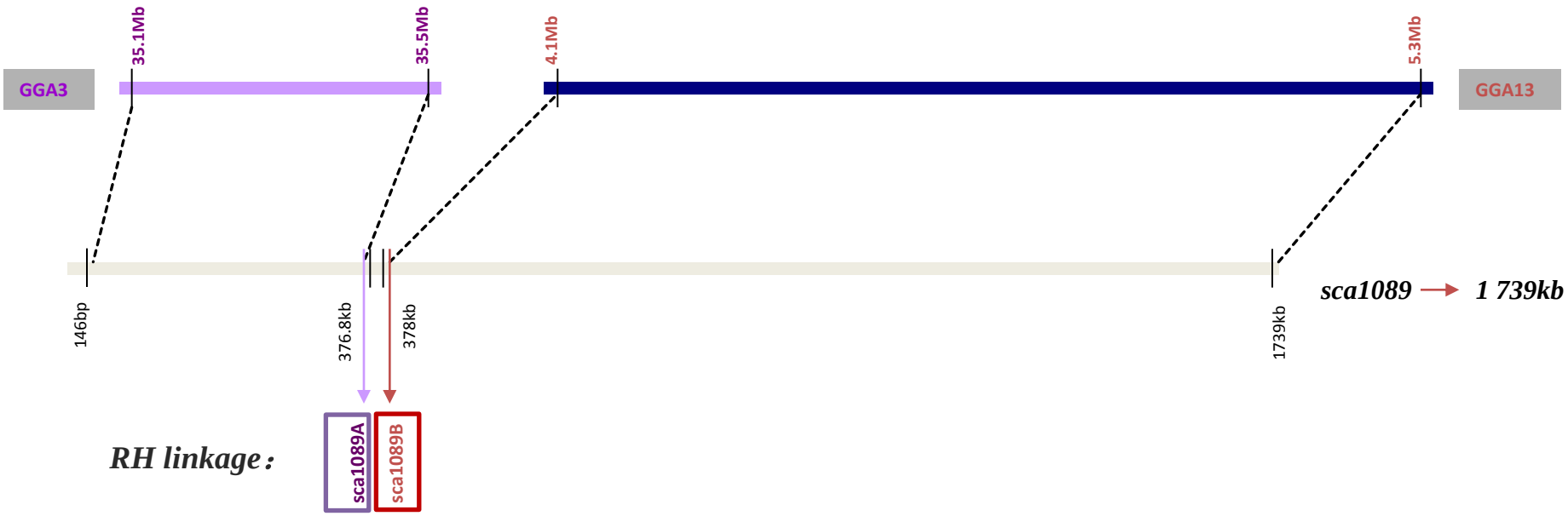


Duck scaffolds aligning to two chicken chromosomes

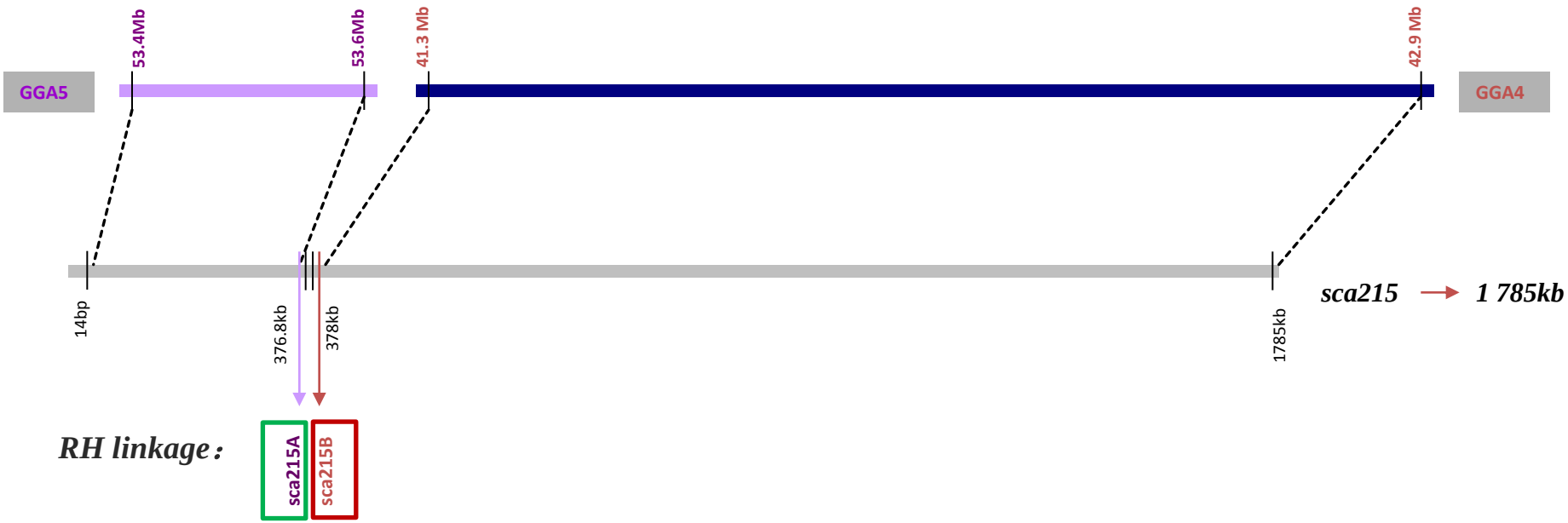




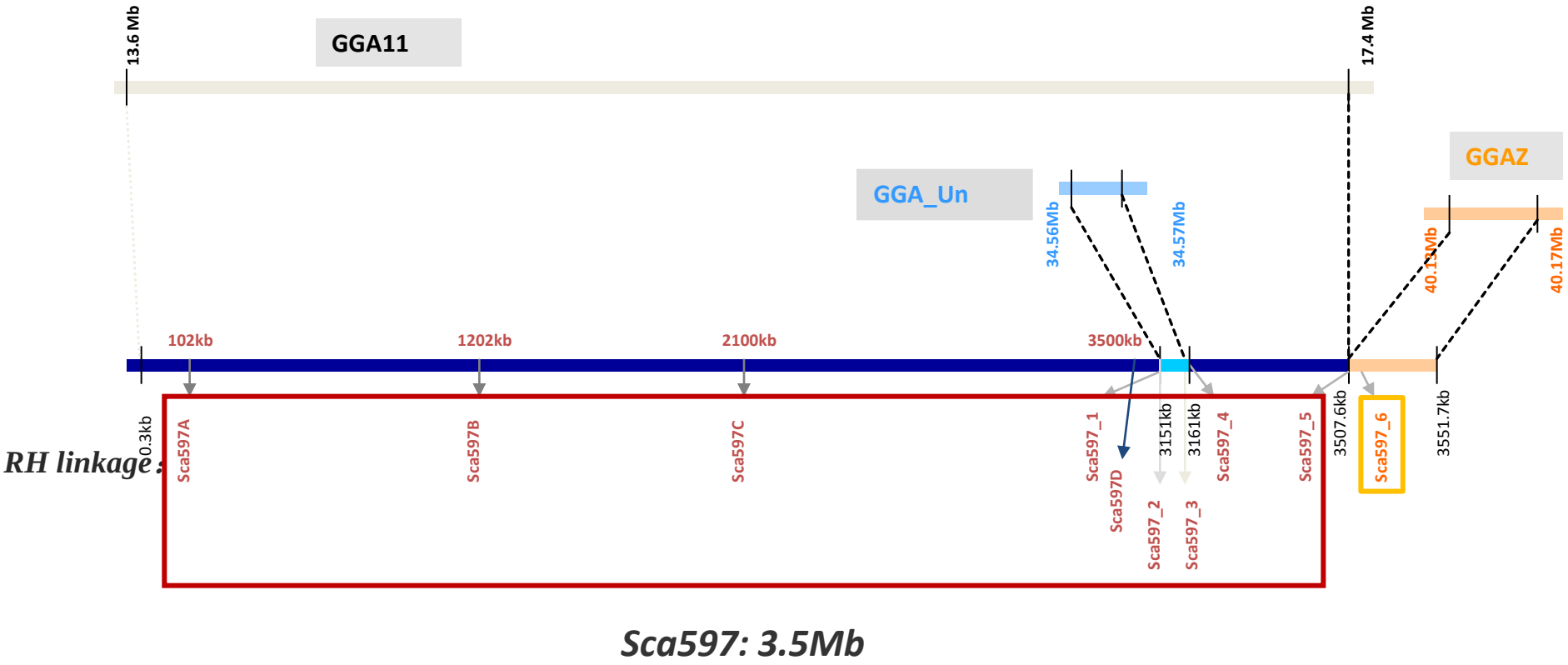
Duck scaffolds aligning to two chicken chromosomes



Duck scaffolds aligning to two chicken chromosomes



Duck scaffolds aligning to two chicken chromosomes



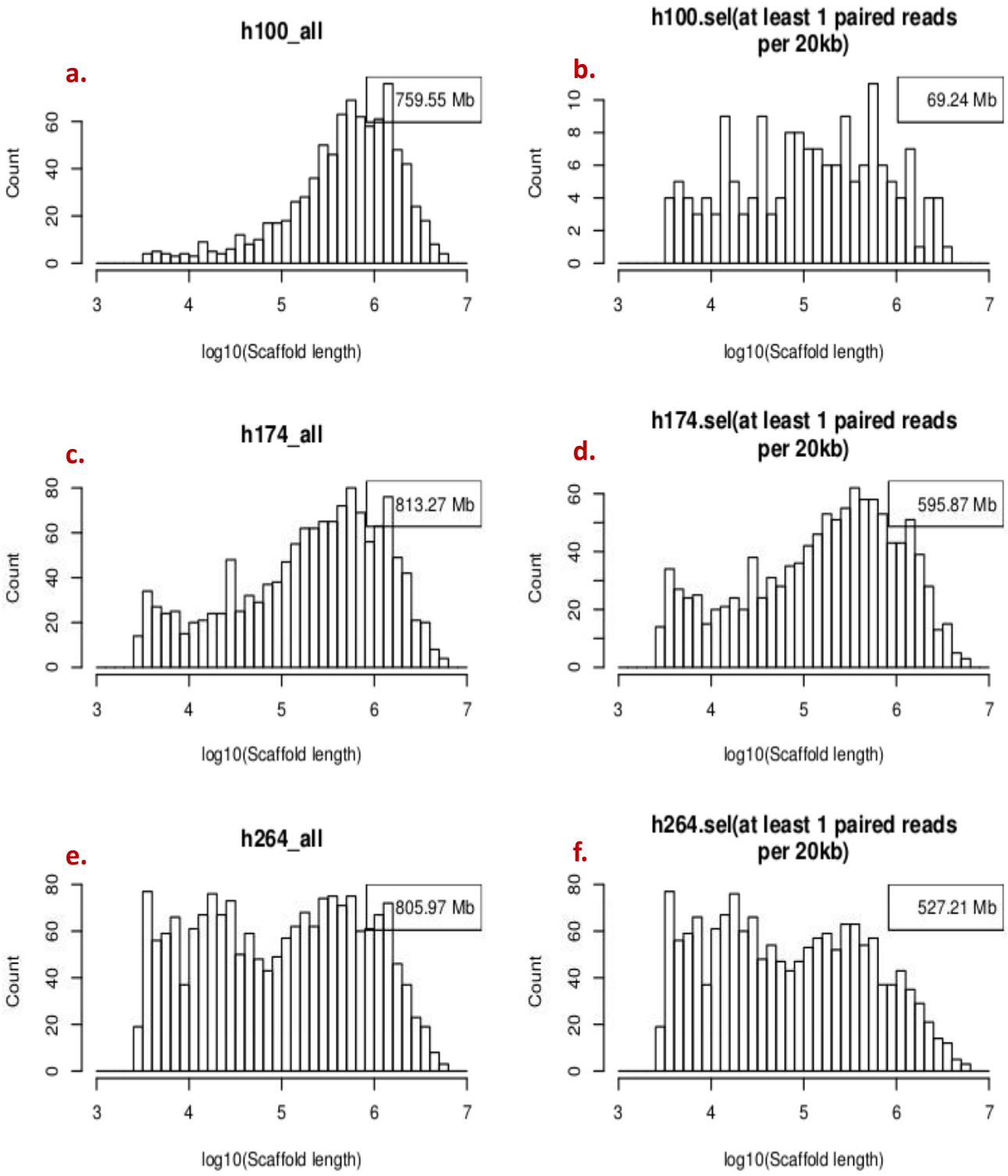
Sca597: 3.5Mb

Scaffold597 was a large scaffold which was detected to be discontinuous as it could be mapped on different chicken chromosome as shown in the graph. The positions of all markers were also indicated. The coordinates on the breakpoints were given as well.

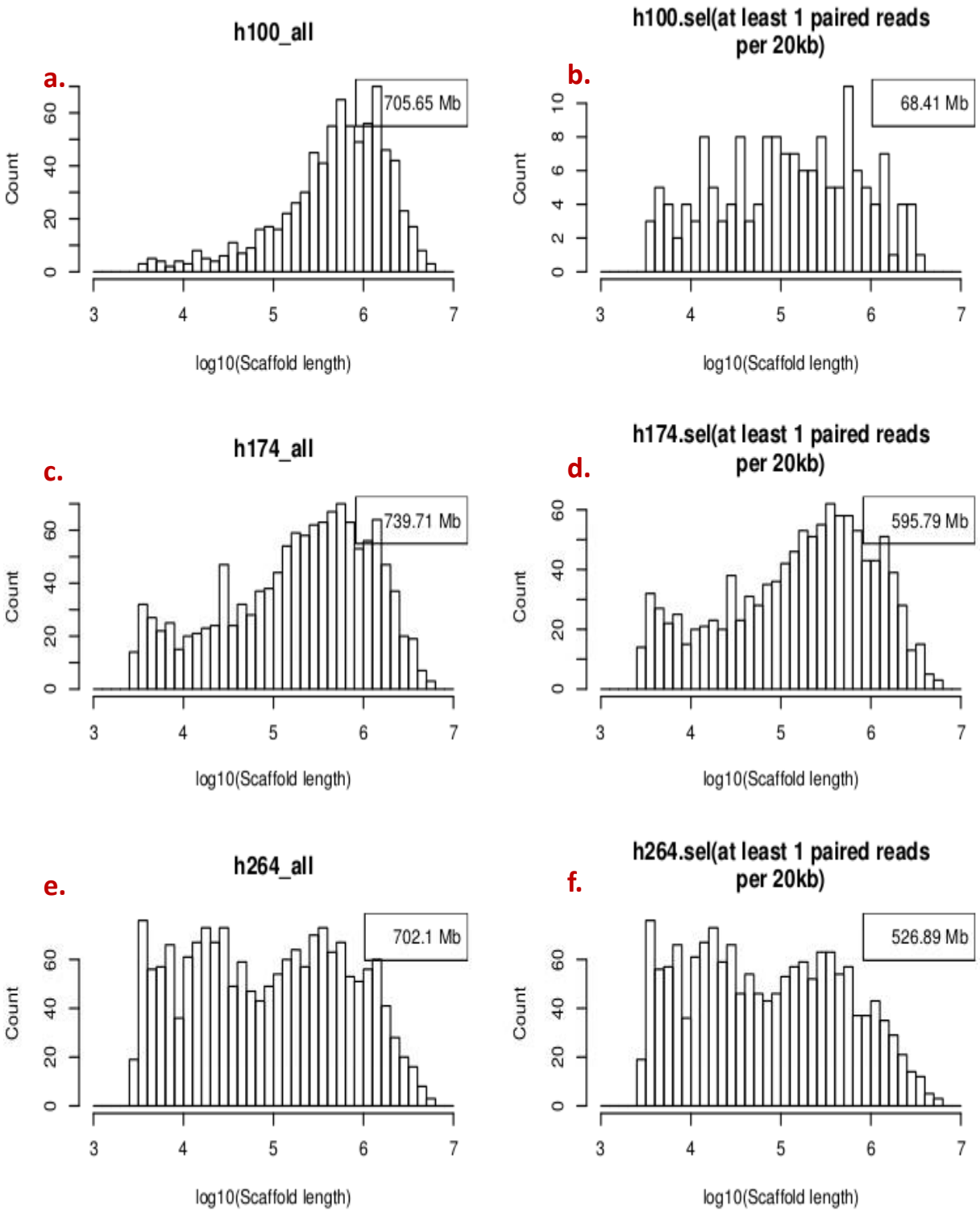
Annex B

Supplementary data to Article in preparation

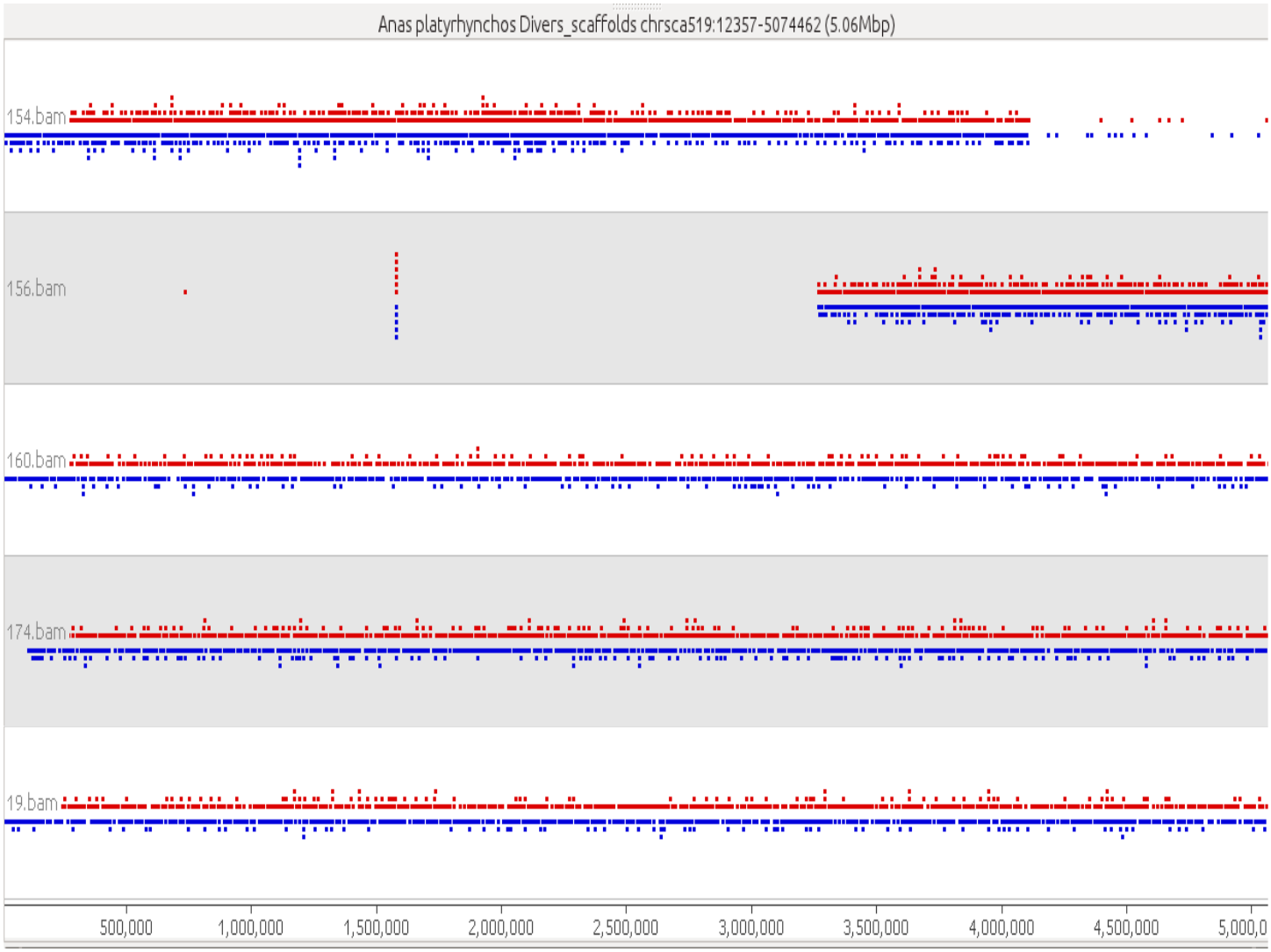
(Chapter IV)



S_Figure1A: **Scaffold length distribution of read-containing scaffolds before removing reads mapped on both species.** Length is shown in logarithm and sum of all mapped scaffolds are shown on the top right of each figure. Three hybrids are represented: h100, h174 and h264. Each row represents a hybrid. e.g (a) distribution of the length of scaffolds having at least one paired read mapped in h100. (b) distribution of the length of scaffolds having at least one paired read per 20kb (or 5 reads/100kb) in h100. Total length in the figure does not reflect total length of duck fragments in the hybrids because scaffolds can be broken.

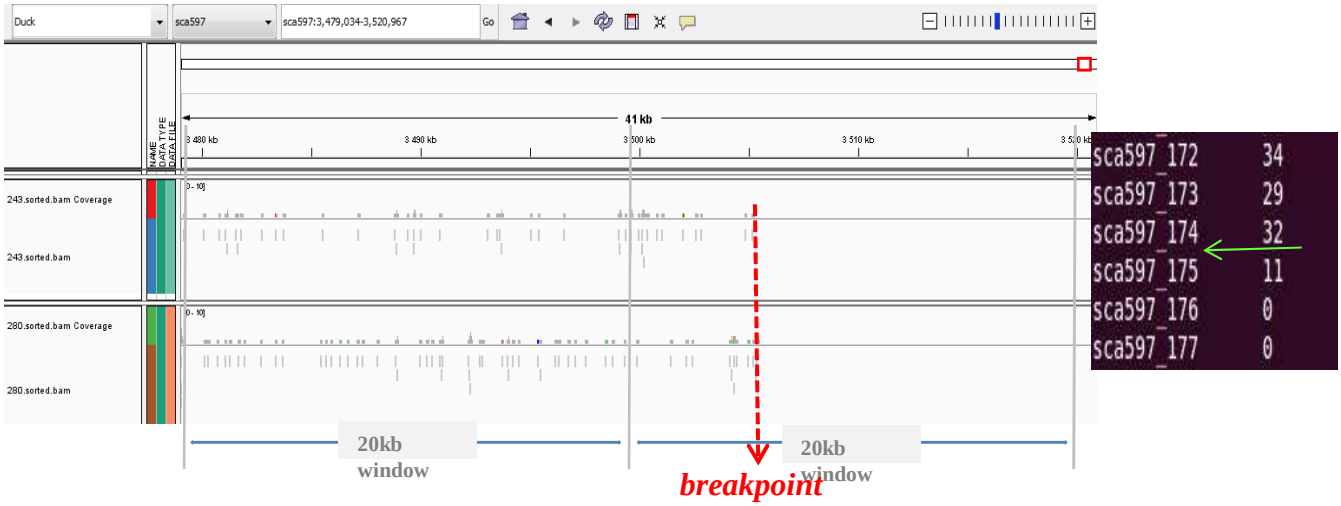


S_Figure1B: : **Scaffold length distribution of read-containing scaffolds after removing reads mapped on both species.** Length is shown in logarithm and sum of all mapped scaffolds are shown on the top right of each figure. Three hybrids represented: h100, h174 and h264. Each row represents a hybrid. e.g (a) distribution of the length of scaffolds having at least one paired read mapped in h100. (b) distribution of the length of scaffolds having at least one paired read per 20kb (or 5 reads/100kb) in h100. Total length in the figure does not reflect total length of duck fragments in the hybrids because scaffolds can be broken.

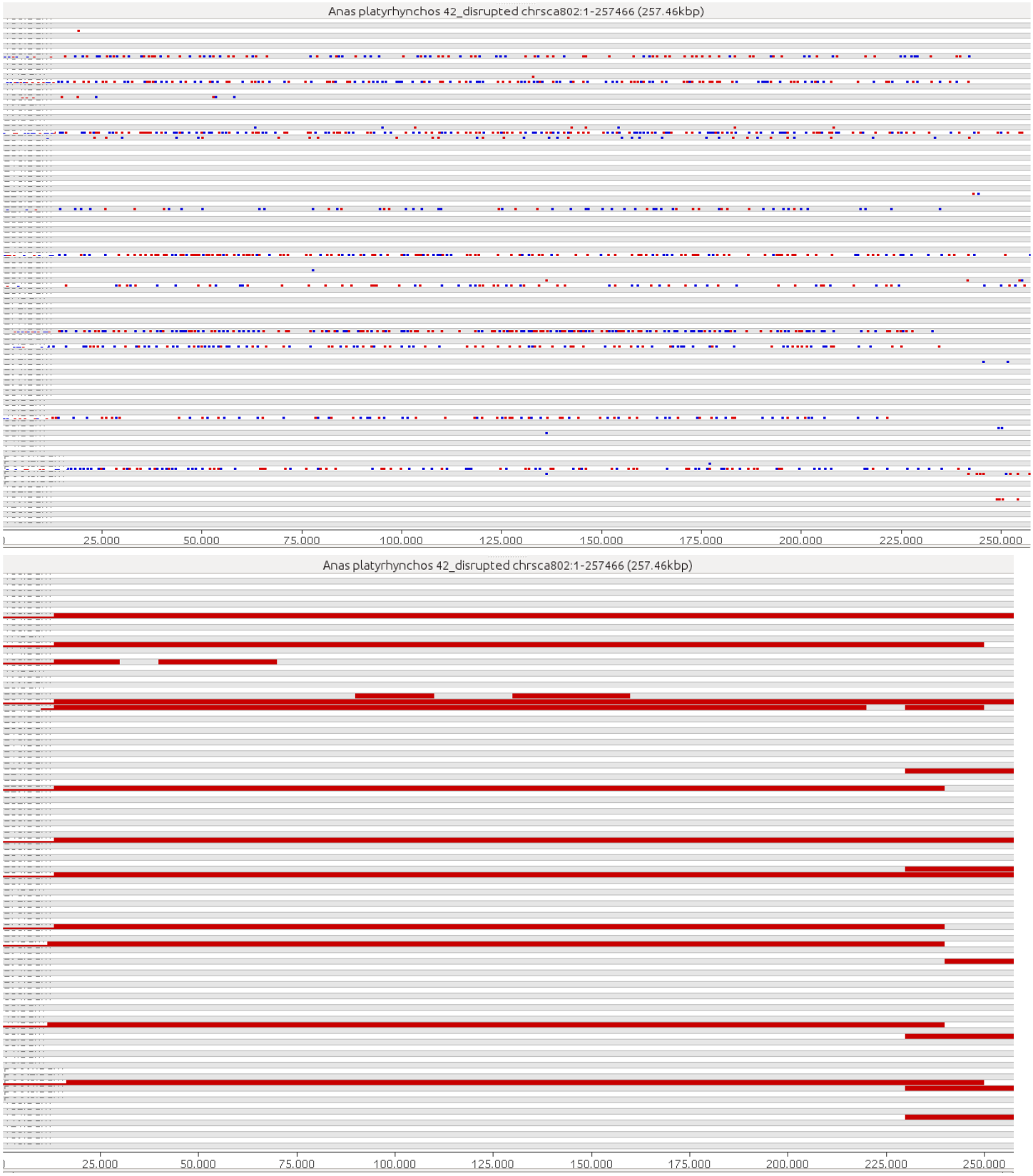


S_Figure 2: Read coverage variation and breakpoints represented by Seqmonk. Sca519 was used as examples for h154, h156, h160, h174 and h19. The total length of sca519 is 5.06Mb. Brown and blue dots in the screenshot represent forward and reverse reads respectively. Read coverage varies amongst hybrids or within hybrids (h154) and a breakage is observed in h156.

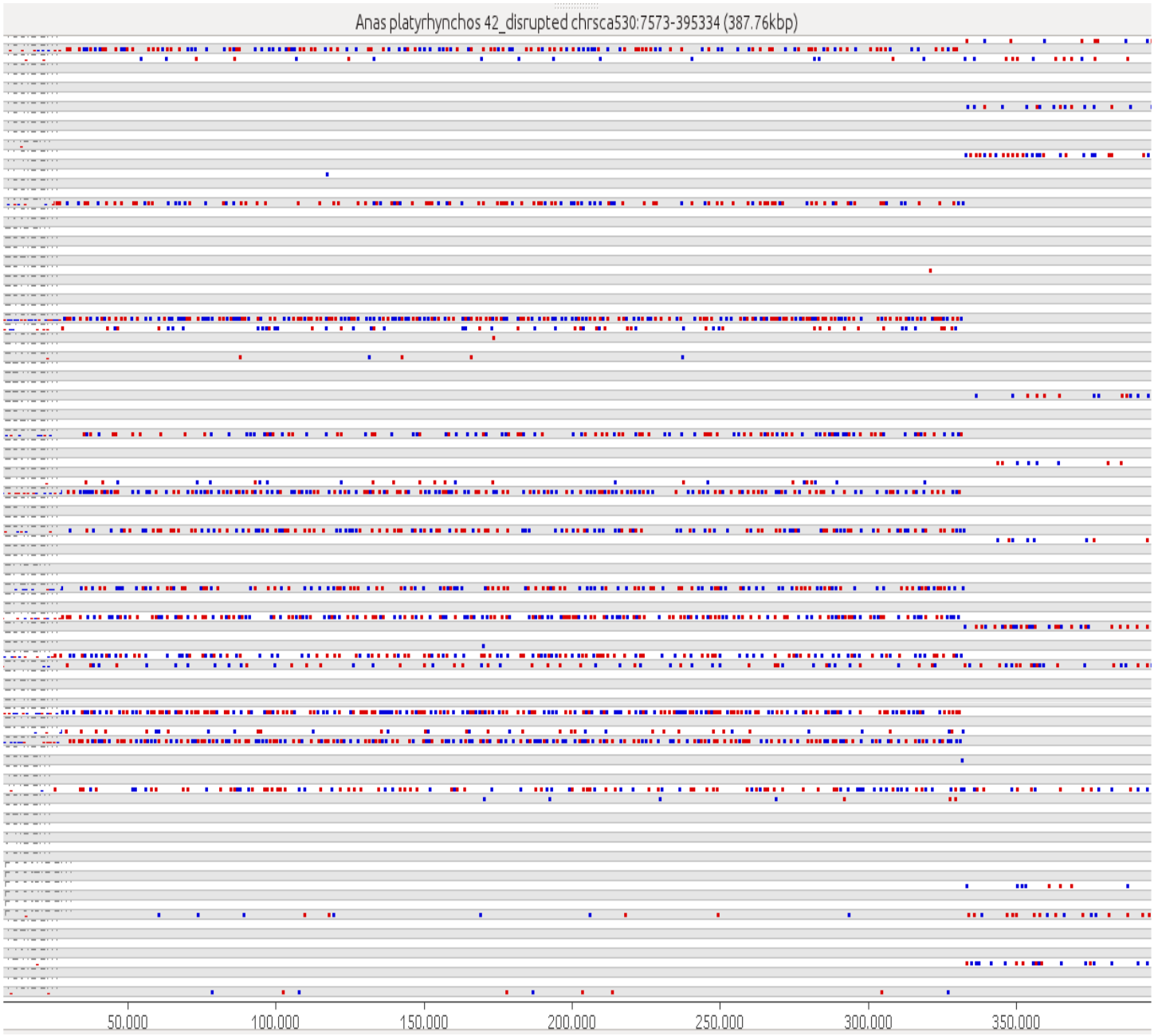
S_Figure 3: **Read coverage variation and breakpoints represented by Seqmonk. Scaffold** sca109 is used as an example in h102, h29, h238, h239 and h295. The total length of sca109 is 3.58Mb. Brown and blue dots in the screenshot represent forward and reverse reads respectively. Hybrid h102 has the highest overall number of sequencing reads and the highest retention fraction among all hybrids showed above, but has the lowest read coverage for this scaffold. Four breakages were observed for this scaffolds in h295.



S_Figure4: Example of false positive calling due to breakpoint imprecision in the CBS segmentation algorithm. When a breakpoint is towards the end of a scaffold, the CBS algorithm can detect it properly, but nevertheless call the small segment as positive. In the example, sca597 is 3554 kb long and is misassembled around position 3506 kb, at 48 kb from the end. The sliding window size used for segmentation was 20kb. Two hybrids h243 and h280 are shown in the figure in IGV and the read counting data for h243 in six windows around the breakpoint, from windows sca_172 to sa_177, is given on the right on a black background. The CBS segmentation suggested that the breakage was between sca597_174 and sca597_175 (green arrow), thus segment 1 goes from the first window (sca597_0) to sca597_174, whereas the last three windows of the scaffold are assigned to segment 2. However, as sca597_175 containing the real breakpoint was assigned to segment 2, the mean read coverage of segment 2 was increased to a point at which it was higher than the threshold for positive calling.

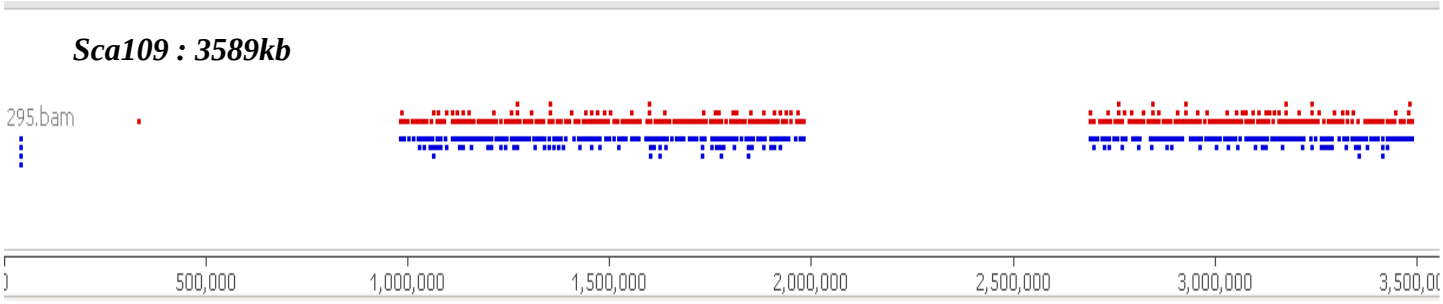


S_Figure 5: **sca802 alignments represented by Seqmonk.** The sequencing reads alignments in all hybrids are shown. Top: brown and blue dots in the screenshot represent forward and reverse reads respectively. Bottom: representation in Seqmonk using a 20kb sliding window with 10kb overlap. The red bars represent presence of a fragment, giving a clearer view on the position of missassembly.



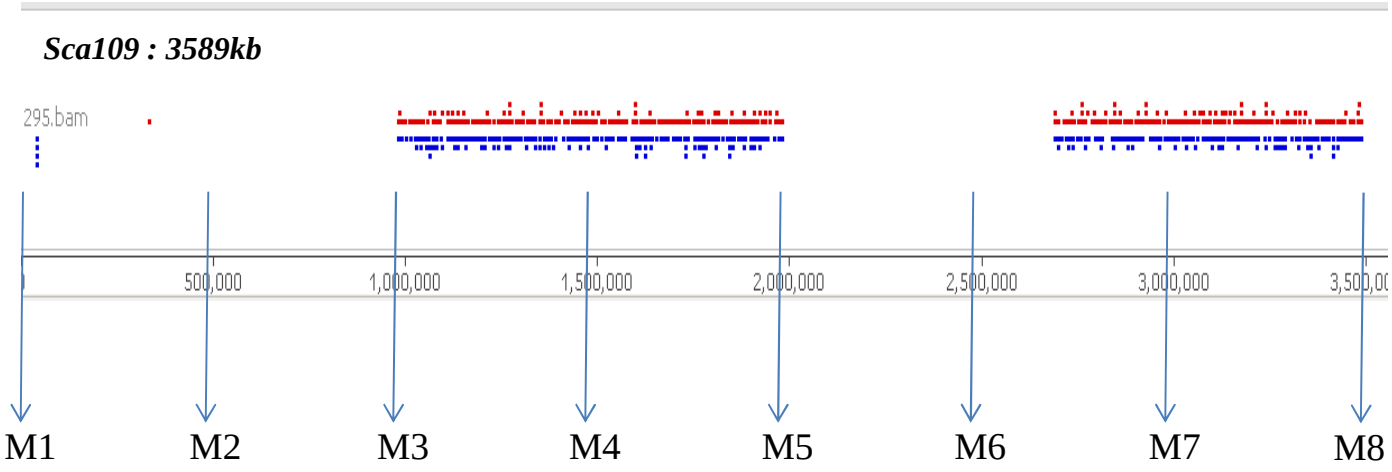
S_Figure 6: **A disrupted scaffolds represented by Seqmonk.** The sequencing reads aligned to sca530 in all the hybrids are shown. Brown and blue dots represent forward and reverse reads respectively. The missassembly appears clearly, with a breakpoint around 330 kb. This missassembly was detected by RH linkage analysis.

a.



Genotyping (CBS): Sca109_0: Absence (A) Sca109_1: Presence (H)

b.



Genotyping (PCR): A-A-H-H-H-A-H-H

S_Figure 7: Comparison of genotyping by sequencing and by PCR (a) Genotyping results for sca109 in h295 by sequencing and CBS segmentation and representation in SeqMonk. In genotyping by sequencing, we only took the status of the scaffold ends into account and therefore the genotyping data suggested this scaffold has one breakage. (b) simulation of a conventional RH mapping by PCR, with markers chosen every 500 kb. Eight markers would be designed from which the genotyping results would reveal three breakages. Therefore the distance between M1 and M8 should be greater than that between sca109_0 and sca109_1.

Supplementary Table1**: sequencing information for hybrids** (reads mapped on both species has been remove

clone	mapped_reads	reads
lane1		88029952
h158	27982	2691514
h17	65140	9654596
h19	105880	8952096
h20	111746	8027348
h27	168016	9190336
h29	189738	8079141
h36	154128	7280104
h38	127318	5516154
h46	83546	7547614
h47	112894	7081940
h50	170205	7701736
h60	89806	6307373
lane2		111424468
h155	186615	6813066
h156	285162	10386194
h159	69194	9325996
h160	185453	9454530
h163	149918	7578305
h164	65770	5561931
h165	73592	12416281
h168	155549	11419990
h170	358602	9440883
h171	233133	10299307
h174	408453	8676271
h185	213955	10051714
lane3		95851511
h188	123250	8175145
h193	162568	11818025
h199	192353	8350941
h200	59274	7136199
h202	136347	7684006
h204	140901	7750970
h206	102059	6941137
h208	120355	8707750
h210	179568	8204979
h213	194444	10997712
h214	133722	8013001
h215	16521	2071646
lane4		118407516
h216	230993	9222456
h219	150408	9958467
h220	222283	8901729
h221	291516	9479779
h223	198123	16724951
h225	227307	17491920
h229	249856	10541728
h232	303194	9706715
h235	163583	9831020
h236	148887	9515739
hpool1	22359	3210700
hpool2	48135	3822312

Supplementary Table1**: sequencing information for hybrids** (reads mapped on both species has been remove

lane5		95219283
h238	79251	8202618
h239	58145	9284622
h243	206098	8160584
h248	225625	7797470
h249	209716	6358376
h250	111362	6702889
h254	177341	7248721
h258	309837	9211296
h259	202397	8731752
h268	60208	6461666
hpool3	110924	8212987
hpool4	256821	8846302
lane6		110501277
h104	217621	9406829
h121	221832	9766156
h231	168297	10854890
h264	340373	10235778
h269	161501	9244696
h270	78229	8356067
h272	129648	8092495
h275	141090	8186134
h276	209030	8125138
h277	230109	10424814
h279	426513	8892600
hpool5	114586	8915680
lane7		130248070
h280	286867	10900124
h289	227109	14506422
h291	77396	5276147
h292	253782	13922418
h293	237284	10656926
h294	218885	10353484
h295	295102	8293463
h297	264267	10972362
h298	301459	9729490
h303	379795	12640921
h304	135739	11119793
hpool6	184516	11876520
lane8		105674153
h62	230683	9672173
h63	209198	9116022
h91	363878	12500409
h94	97109	9536064
h96	241833	9997706
h100	40426	6457586
h102	280435	10051377
h119	94337	8902129
h133	323369	9004227
h139	147830	6859680
h140	139225	6543365
h154	186448	7033415
lane9		31145549
h150	100793	4346000
h201	176601	6656681
h207	155906	6793978
h290	169512	13348890

NIGMS HUMAN/RODENT SOMATIC CELL HYBRID MAPPING PANEL #1

PERCENTAGE OF CELLS WITH HUMAN CHROMOSOMES

CHROMOSOME#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	Y
HYBRID/DNA																								
GM/NA09925	74	24	0	74	76	60	82	78	0	0	4	68	6	86	78	14	98	96	46	84	0	76	0	0
GM/NA09926	69	75	75	65	2	88	85	69	0	68	0	2	77	73	93	2	81	75	84	96	2	4	2	0
GM/NA09927	69	83	75	77	0	93	79	73	0	82	0	0	77	79	90	0	81	73	87	89	0	0	0	0
GM/NA09928	0	84	58	0	48	32	0	66	0	2	0	0	4	76	92	0	98	0	28	0	70	82	0	78
GM/NA09929	0	0	61	59	0	43	2	49	0	0	33	49	0	59	2	0	96	0	2	31	0	0	2	0
GM/NA09930A	0	34	62	4	12	0	26	4	0	0	6	22	56	82	12	0	86	78	0	22	82	76	6	8
GM/NA09931	0	0	0	0	26	0	78	0	0	46	0	64	0	100	0	0	100	0	0	78	90	0	0	14
GM/NA09932	0	0	0	68	86	46	0	80	0	2	28	26	0	0	0	0	96	0	2	0	92	0	0	0
GM/NA09933	50	0	84	16	54	76	92	54	0	6	0	50	84	78	92	0	88	70	80	32	94	88	0	32
GM/NA09934	0	50	0	0	83	79	4	87	0	0	77	87	0	2	89	0	90	89	0	91	89	2	0	0
GM/NA09935A	0	0	52	10	28	12	0	0	0	8	0	22	74	72	0	0	93	59	0	9	91	71	0	0
GM/NA09936	0	0	0	18	0	46	70	10	0	16	34	0	2	88	2	0	100	0	44	24	0	18	0	0
GM/NA09937	0	0	54	38	0	62	54	70	0	4	0	42	0	70	60	0	96	66	0	0	0	0	0	0
GM/NA09938	0	0	2	88	60	88	86	4	0	0	36	92	0	80	4	0	92	0	4	80	76	60	0	2
GM/NA09940	0	0	46	0	0	0	84	62	0	0	0	0	0	0	62	0	100	0	0	0	0	0	0	0
GM/NA10324	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	90	0
GM/NA10567	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	98	0	0	0	0	0	0	0	0
GM/NA10611	0	0	0	0	0	0	0	0	69	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

The above percentages are the averages of the results obtained for the cytogenetic analysis of a minimum of 25 cells examined both at first passage and at final harvest.

S_Table 2: **chromosome counting human somatic cells**. Each Column represents the percentage of cell harbouring corresponding human chromosome. Each row was results in a single hybrid. It is clear that the human somatic hybrid cell lines are a mixture of cells containing different human chromosomes (A.Vignal personnel communication)

Supplementary Table3: **example of the CBS segmentation couldn't detect breakage**

hybrid	scaffold	length (bp)	window	read counts	window size
# h163	sca2049	96325	sca2049_0	0	20000
			sca2049_1	0	20000
			sca2049_2	2	20000
			sca2049_3	5	20000
			sca2049_4	1	16325
# h188	sca2403	221726	sca2403_0	0	20000
			sca2403_1	0	20000
			sca2403_2	0	20000
			sca2403_3	0	20000
			sca2403_4	0	20000
			sca2403_5	0	20000
			sca2403_6	0	20000
			sca2403_7	2	20000
			sca2403_8	5	20000
			sca2403_9	3	20000
			sca2403_10	6	20000
			sca2403_11	1	1726
# h188	sca638	254101	sca638_0	6	20000
			sca638_1	5	20000
			sca638_2	6	20000
			sca638_3	5	20000
			sca638_4	12	20000
			sca638_5	8	20000
			sca638_6	16	20000
			sca638_7	5	20000
			sca638_8	6	20000
			sca638_9	6	20000
			sca638_10	0	20000
			sca638_11	0	20000
			sca638_12	0	14101
# h193	sca1096	196431	sca1096_0	18	20000
			sca1096_1	18	20000
			sca1096_2	31	20000
			sca1096_3	19	20000
			sca1096_4	28	20000
			sca1096_5	23	20000
			sca1096_6	21	20000
			sca1096_7	0	20000
			sca1096_8	0	20000
			sca1096_9	0	16431

Supplementary Table4: **the potential breakpoint region detected by SeqMonk** for disrupted scaffolds (NA: not applicable)

scaffold	disrupted(yes or no)	misassembly point
sca102	yes	360-385kb
sca1083	yes	62-74kb
sca1197	yes	720-735kb
sca1375	yes	1005-1010kb
sca1517	yes	558-562kb
sca180	no	NA
sca1893	yes	220-221kb
sca193	yes	706-707kb
sca2049	yes	47,4-48kb
sca215	yes	238-239kb
sca227	yes	1216-1218kb
sca245	yes	56-61kb
sca246	yes	619-622kb
sca279	yes	30-50kb
sca286	yes	756-757kb
sca3008	yes	21-23kb
sca316	yes	208-212kb
sca3271	yes	141-142kb
sca354	yes	1372-1376kb
sca365	yes	364-370kb
sca367	yes	308-310kb
sca398	?	NA
sca420	yes	492-493kb
sca458	yes	1739-1740kb
sca481	yes	188-190kb
sca504	yes	3977-3979kb
sca513	yes	675-678kb
sca530	yes	332-333kb
sca597	yes	3507-3508kb
sca629	yes	174-175kb
sca649	no	NA
sca676	yes	198-199kb
sca77	yes	99-102 kb
sca802	?	NA
sca810	yes	332-326kb
sca851	yes	1727-1727,4kb
sca868	yes	636-638kb
sca870	yes	446-448kb
sca881	yes	110-111kb
sca945	yes	139-141kb
sca956	yes	945-950kb
sca530	yes	326-328kb

Supplementary Table5: **the evolutionary breakpoint region analysis for the three chromosomes**.

scaffold	start	end	AT	GC	N (gap)	MaskedRepeat(family)	Repeat_start	Repeat_end
APL22								
sca246	642000	658535	51,58%	45,37%	3,05%	NA		
sca246	915760	931501	44,63%	51,87%	3,50%	NA		
sca246	1000919	1005919	44,73%	53,59%	1,68%	NA		
sca3327	1	5000	50,12%	49,88%	0,00%	NA		
sca3327	36189	41819	48,49%	51,39%	0,12%	NA		
sca1197	1	5000	36,60%	41,80%	21,60%	NA		
sca1197	593946	598946	48,11%	47,03%	4,86%	NA		
sca871	327216	343394	56,70%	43,30%	0,00%	GGLTR5B(LTR/ERVL)	328236	328611
sca2665	1	111922	37,79%	56,12%	6,09%	GGLTR5A(LTR/ERVL)	13890	14167
sca1885	1	298994	48,85%	43,45%	7,70%	CR1-Y2_Aves	140324	140750
sca871	341000	346500	54,44%	45,56%	0,00%	NA		
sca871	32400	329216	51,50%	42,61%	5,88%	NA		
APL12								
sca736	550000	55159	55,83%	44,17%	0,00%	NA		
sca736	572000	577000	65,13%	33,59%	1,28%	NA		
sca903	363000	368878	62,54%	37,46%	0,00%	NA		
sca903	383000	388494	60,96%	32,68%	6,35%	NA		
sca743	1	10000	54,07%	37,74%	8,19%	CR1-F(LINE/CR1)	3483	4556
sca469	1	39042	21,69%	31,99%	46,32%	NA		
sca469	39403	44042	44,63%	43,79%	11,57%	NA		
sca769	1	6000	62,62%	37,38%	0,00%	NA		
sca3421	1	22321	23,29%	43,04%	33,68%	NA		
sca2156	1	10000	56,42%	43,58%	0,00%	GGLTR5A(LTR/ERVL)	3631	3870
sca1434	1180000	1233631	48,55%	46,66%	4,79%	NA		
sca26	1	6000	48,30%	51,70%	0,00%	NA		
APL2								
sca74	2382008	2387008	62,35%	37,65%	0,00%	GGLTR8B(LTR/ERVL)	2386499	2386957
sca74	2374046	2382008	60,63%	33,48%	5,89%	NA		
sca1034	504218	509218	63,73%	36,27%	0,00%	NA		
sca1034	495371	5000371	61,05%	38,95%	0,00%	NA		
sca1034	500372	504217	63,49%	36,51%	0,00%	NA		
sca1497	1	5000	57,86%	42,14%	0,00%	CR1-E(LINE/CR1)	937	1545
sca1497	50000	54938	55,40%	35,45%	9,15%	CR1-F2(LINE/CR1)	500251	501465
sca488	309030	314030	65,23%	34,77%	0,00%	NA		
sca1190	1	5000	65,18%	34,82%	0,00%	NA		
sca1190	39388	44388	54,39%	45,61%	0,00%	CR1-F2(LINE/CR1)	43569	44067
sca258	1	5000	59,68%	40,32%	0,00%	NA		
sca22	3124000	3129084	64,39%	35,61%	0,00%	NA		
sca5	1	5000	63,72%	36,28%	0,00%	CR1-X2(LINE/CR1)	1104	1463
sca5	28924	33924	55,53%	44,47%	0,00%	CR1-C4(LINE/CR1)	32419	32615
sca280	1	6000	58,92%	41,08%	0,00%	NA		
sca529	891000	896469	58,63%	41,37%	0,00%	NA		
sca529	900793	905793	59,07%	40,93%	0,00%	CR1-C4(LINE/CR1)	901402	901847
sca529	900793	905793	59,07%	40,93%	0,00%	CR1-Y(LINE/CR1)	901846	902946
sca2592	188000	194255	57,51%	42,49%	0,00%	CR1-Y(LINE/CR1)	192518	194256
sca616	22082	28082	63,54%	36,46%	0,00%	CR1-X2(LINE/CR1)	26518	26867
sca616	22082	28082	63,54%	36,46%	0,00%	CR1-C4(LINE/CR1)	27007	27186
sca616	14665	19665	62,09%	37,91%	0,00%	NA		
sca1008	125157	130157	66,13%	33,87%	0,00%	NA		