# The Iccare web server: an attempt to merge sequence and mapping information for plant and animal species.

C. Müller, M. Denis, Laurent Gentzbittel, Thomas Faraut

*Nucleic Acids Research, 2004, Vol. 32, Web Server issue* ***W429–W434***
*DOI: 10.1093/nar/gkh460*

# The Iccare web server: an attempt to merge sequence and mapping information for plant and animal species

**Cédric Muller[1], Mathieu Denis, Laurent Gentzbittel[1] and Thomas Faraut***

INRA, Laboratoire de génétique cellulaire and [1]INP-ENSAT, Laboratoire de biotechnologies et d'amélioration des plantes, Castanet Tolosan 31326, France

Received February 16, 2004; Revised and Accepted April 26, 2004

## ABSTRACT

**The Iccare web server, http://genopole.toulouse.inra.fr/bioinfo/Iccare, provides a simple yet efficient tool for crude EST (expressed sequence tag) annotation specifically dedicated to comparative mapping approaches. Iccare uses all the EST and mRNA sequences from public databases for an organism of interest (query species) and compares them to all the transcripts of one reference organism (*Homo sapiens* or *Arabidopsis thaliana*). The results are displayed according to the location of the genes on the chromosomes of the reference organism. Gene structure information and sequence similarities are combined in a graphical representation in order to pinpoint the nature of the transcript query sequence. The user can subsequently design primers or probes for the purpose of physical or genetic mapping. In addition to the query organisms already available in Iccare, users can perform a tailor-made search with their own sequences against the animal or plant reference organism genes.**

## INTRODUCTION

Comparative analysis has always been central to biology. With the advent of the DNA revolution, the development of computer tools for sequence comparison has been essential to apply a comparative approach at the protein or DNA level. The success of this approach is attested by the extensive use of the world-famous BLAST programs [(1), W. Gish, http://blast.wustl.edu]. In the field of gene mapping, the comparative approach takes advantage of genome conservation. Indeed, while at the molecular level genes evolve by accumulated point mutations, at the genome scale chromosomes evolve by inter- and intra-chromosomal rearrangements of large segments within which the gene content and order may remain unaltered (2–4). Comparative mapping studies have confirmed the local conservation of gene repertoire and order, also called microsynteny conservation, not only between mammals (5–7) but also between mammalian species and more distant animals such as pufferfish and chicken (8–10). For plants, this conservation is clearly revealed between species of the same family [Poaceae (11–13) or Brassicaceae (14–16)] and also between *Arabidopsis thaliana* (Brassicaceae) and species of other families [Fabaceae (17–19), Poaceae (20,21) and Solanaceae (22,23)]. Without underestimating the confusing effect of local micro-rearrangements, the synteny conservation enables, at least to a certain extent, the transfer of mapping information from one species to another and many genome species are now studied using the human (24–28) or the *A.thaliana* (29–33) genome dense maps to accelerate the process of mapping.

With the emergence of high-throughput sequencing projects, which generate both EST (expressed sequence tag) and genomic DNA, an impressive amount of sequence data from many species is available in public databases. Whereas much effort has been devoted to developing computer tools for sequence assembly, annotation (34,35) and their graphical display (36,37), the process of gathering sequence data in a chromosomal region of interest for a given genome is still a time-consuming and awkward task. The main comparative mapping studies are usually oriented towards local chromosomal investigations for Quantitative Trait Loci (QTL) mapping or fine mapping purposes. In the case of a fine mapping approach in a mammalian species, the process starts by identifying, using state-of-the-art comparative maps, the corresponding region in the human genome. The gene sequences located in the so-called homologous region are used to identify available orthologous ESTs for the genome under study. These EST sequences are subsequently used to design primers for mapping experiments.

To facilitate and accelerate the exploitation of public sequence data available for different plant and animal species—the query species—we propose organizing them according to their sequence similarities to the genes of a completely sequenced organism—reference organism (*Homo sapiens* for animals and *A.thaliana* for plants). This is the underlying

*To whom correspondence should be addressed. Tel: +33 05 61 28 54 31; Fax: +33 0 5 61 28 53 08; Email: Thomas.Faraut@toulouse.inra.fr

principle of the tool Iccare, which stands for 'Interspecific Comparative Clustering and Annotation foR Est', available at http://genopole.toulouse.inra.fr/bioinfo/Iccare. Using this tool, sequence similarities results are available through the chromosomal maps of the reference organism. Furthermore, based on the gene splicing information for the reference organism, Iccare enables the prediction of splicing site positions on the query species' sequences. In addition to publicly available data, the user can organize new sequences according to the same scheme.

## MATERIALS AND METHODS

### Datasets

For the human genome, the UniGene clusters were used to define the reference organism gene catalog [(38), http://www.ncbi.nlm.nih.gov]. The set of unique sequences is used to define the reference transcript sequence for each gene (see the README file at ftp://ftp.ncbi.nih.gov/repository/UniGene). The mapping information as well as the gene structure information is defined according to the annotation of the human genome provided by Ensembl [(37), http://www.ensembl.org]. For the *Arabidopsis* genome, the gene sequences (26, 637), gene information and mapping information were defined according to the Munich Information Center for Protein Sequences [(39), http://mips.gsf.de]. For *Arabidopsis* transcript sequences, only the translated regions of the predicted mRNA were used. For the query animal or plant species (Table 1), all EST and full-length mRNA sequences have been downloaded using the SRS retrieval software on the Infobiogen server (http://www.infobiogen.fr). Additional organisms can be added on request to the authors.

### Personal inputs

Personal sequences can be also submitted to Iccare. The input required by Iccare is a sequence or a set of sequences in FASTA format.

### Software description

All sequences are screened for vectors and masked for known repeats with RepeatMasker (UniVec at ftp://ftp.ncbi.nih.gov/pub/UniVec; A. Smit and P. Green, RepeatMasker at http://ftp.genome.washington.edu/RM/RepeatMasker.html). The masked sequences are subsequently compared to all the transcript sequences of the reference organism with the blastn option of the BLAST program (1). The results are filtered according to the expected value. This value is normalized in order to fit a standardized expected (*E*-)value of a comparison with a database of one million residues. Only similarities—or more precisely, using the terminology of BLAST, highest scoring pairs (HSPs)—with an *E*-value $<10^{-5}$ are kept for further consideration (the complete BLAST output is also available). Finally, Iccare compiles the BLAST results and the mapping information for the reference organism and formats these results for the web site application.

### Website dynamic script

Programming was done in Perl using the CGI library, the GD library and the EMBOSS package (40).

**Table 1.** Available organisms on Iccare

| Animals | Plants |
|---|---|
| *Bos Taurus* | *Brassica napus* |
| *Canis familliaris* | *Brassica rapa* |
| *Capra hircus* | *Chlamydomonas reinhardtii* |
| *Equus caballus* | *Helianthus annuus* |
| *Gallus gallus* | *Medicago truncatula* |
| *Oryctolagus cuniculus* | *Physcomitrella patens* |
| *Ovis aries* | *Oryza sativa* |
| | *Zea mays* |

## ICCARE WEBSITE

The taxonomic tree on the Iccare homepage presents the different query organisms that can be selected for analysis. These organisms are compared to an appropriate reference organism (currently limited to *H.sapiens* for animal species and *A.thaliana* for plant species). The results are presented on dynamic web pages. A tutorial and help pages are also available.

After having selected the query organism and a model chromosome, the first result page (Figure 1) presents a graphical representation of the chromosome of the reference organism together with a visualization of the gene distribution with particular emphasis on genes showing sequence similarities to the sequences of the query organism. Moreover, a BLAST link allows a blatsn or tblastx search for each query organism sequence against the gene catalog of the reference organism. This gives an idea of the gene family potentially associated with the putative ortholog. The 'alignment' link provides a graphical representation of the gene structure of the reference organism (translated region, exon splicing positions and intron size) and the query sequence similarity with this gene (Figure 2A). The alignment—global or local—between the reference and the query sequences is also available (Figure 2B). As the gene structure is known to be very well conserved (41), this information makes it possible to predict the structure (exon borders) of the query sequence. Various links have been devised for the sake of practicality: Primer3 for primer design (42) and Overgo Maker 40 (http://www.genome.wustl.edu).

## RESULTS AND DISCUSSION

The Iccare tool has proven to be practical and efficient for various animal and plant studies. In animals, Iccare has been used in the context of comparative mapping studies between human and pig (43–45), chicken (46) and bovine (47). More recently, in plants, the use of Iccare has been of great help for the construction of sunflower genetic maps or physical maps. The possibility of inferring the putative exon splicing sites increases the success rate for designing polymorphic markers from 15–20% by random primer design to 65% (Delphine Samson, GenoPlante, personal communication). In addition, the identification of conserved EST regions facilitates the design of overgos for physical mapping.

Indeed, with the availability of state-of-the-art comparative maps for bovine and human species, it is straightforward to identify all the bovine transcribed sequences available in the public databases that have sequence similarities to the human genes in the region of interest. Moreover, the graphical display

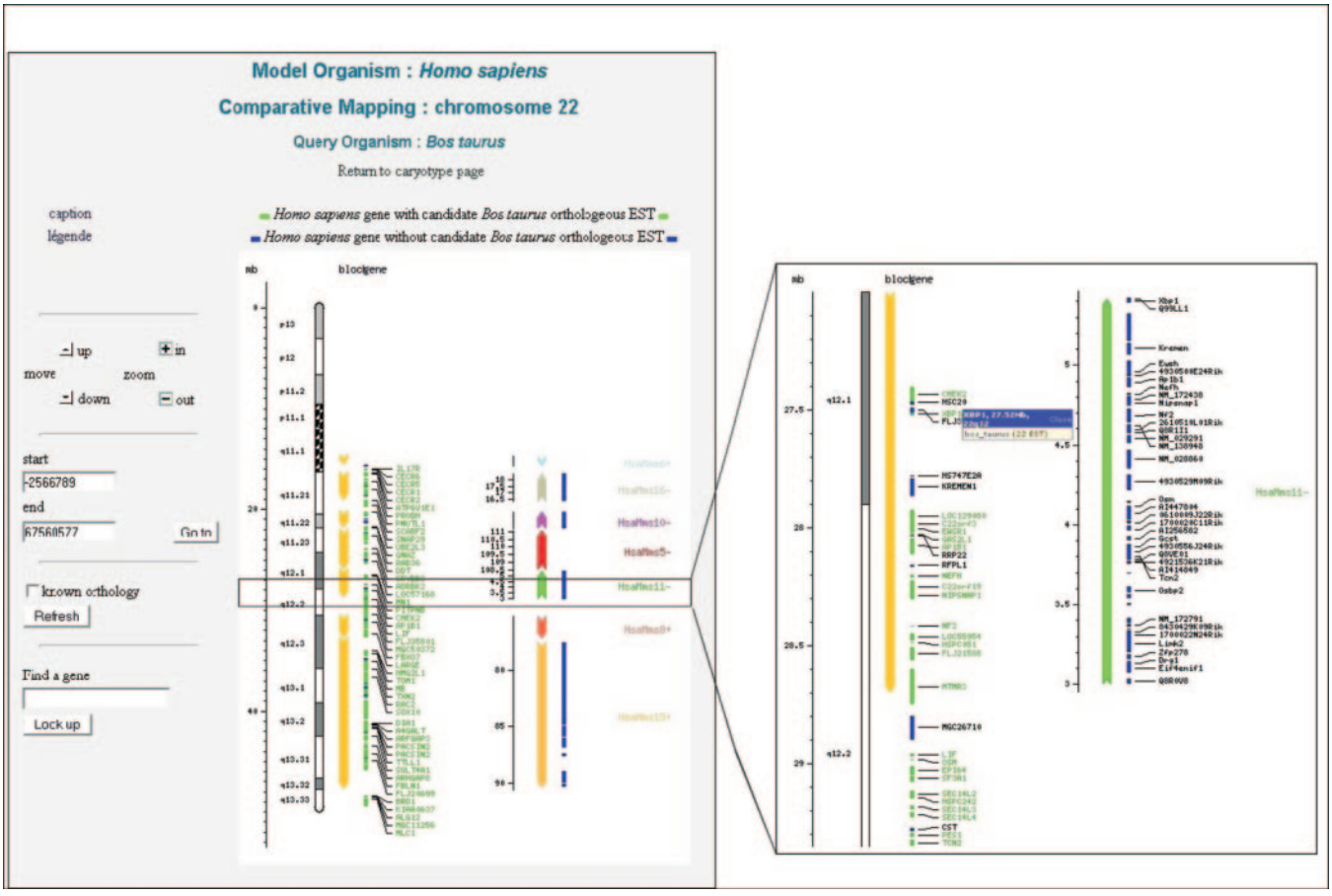


**Figure 1.** Global chromosome synteny. Example of human chromosome 22 gene map for the query species *Bos taurus* at the chromosome scale on the left and at a higher resolution on the right. From left to right, the human chromosome image displays map locations in Mb and a cytogenetic map with alternating dark and light bands, with a dark-and-white check pattern for the centromeric region. The dark-yellow arrowed rectangles display the conserved syntenic segments shared with the mouse genome. To the right, the corresponding conserved syntenic segments of the mouse genome are displayed in colors corresponding to the mouse chromosome. The buttons on the left enable the user to move to different regions at different levels of resolution. The genes colored in green correspond to human genes for which sequence similarities have been observed to bovine sequences. Human genes without significant similarities to bovine sequences are shown in blue. The mouseover on a gene provides a link that gives access to detailed information about the alignments (see Figure 2).

of sequence similarities allows the addition of pertinent information to the sole information of the *E*-value associated with the score as given by alignment software. A local similarity restricted to the coding region of the human transcript is in good agreement with the scheme of sequence evolution. Furthermore, additional features of the alignment such as phase conservation and mismatches occurring essentially on the third codon position correspond to qualitative characterization of the alignment that are not taken into account by the statistical model of sequence similarities. Finally, additional sequence comparisons can be carried out by the user in order to clarify the relationship between the transcripts of the species of interest and the reference organism transcripts. This might pinpoint for instance an alternative nature of a particular EST.

It has to be pointed out that the distinction between orthologous and paralogous genes on the basis of sequence similarities is a difficult issue. Other programs are specifically dedicated to this purpose (48–50). This problem should therefore be kept in mind when using the simple sequence similarities results provided by Iccare. The question is even more problematic in plants where the genome seems to have evolved by several rounds of polyploidy and/or chromosomal duplication (51). Nevertheless, the simple procedure proposed by our software makes it possible to assume membership of the query sequence to a gene family, which enables further consideration using specific software.

## CONCLUSION

With the accumulation of massive amounts of sequence data for many organisms, the problem arises of proposing efficient computer tools to facilitate access to data. A trade-off has to be made between general-purpose databases where all the information is available but difficult to exploit and completely automated systems proposing annotated EST sequences clustered and associated with consensus sequences which are tentatively the reconstructed mRNA. Although both systems are necessary for different purposes, semi-automated systems, such as Iccare, can be of great help for the exploitation of sequence data.


**Figure 2.** Gene structure representation and local alignment. (**A**) The *Arabidopsis* gene structure of At4g03280 and *Helianthus annuus* sequences with homologous regions with this gene are shown. The top graphical box contains gene structure information related to the *A.thaliana* gene—intron/exon structure of genomic DNA. The green boxes represent the exons, which are concatenated on the second line, mimicking the transcription process. The numbers on top of the first line correspond to intron size. The third line recalls the transcript sequence with a particular emphasis on the translated region, shown in dark blue. The bottom graphical box contains *H.annuus* sequences and representation of local similarities to the *Arabidopsis* gene At4g03280; information for each query sequence is located to the left (GenBank ID, sequence size and BLASTN *E*-value) and the sequence representation is on the right: the black line symbolizes the sequence. Green boxes correspond to sequence similarities with the *Arabidopsis* gene on the same strand, while yellow boxes correspond to similarities in a reverse-complement manner. Sequence similarities to additional query species are also available ('plus' buttons). (**B**) The display associated with the result of a local alignment between *Arabidopsis* gene At4g03280 and the *H.annuus* EST sequence CD847222. BLASTN information (score, *E*-value, identities, gaps and strand) is placed on the top. The query nucleic sequence is represented in black and the *Arabidopsis* nucleic sequence is represented in alternate blue and mauve letters corresponding to the different exons.

## ACKNOWLEDGEMENTS

We thank the many researchers who contributed to Iccare by their useful suggestions. We also thank the Génopôle Toulouse Midi-Pyrénées (France), and especially David Allouche, for providing the computer and computer administration resources for data processing and the web service.

## REFERENCES

1. Altschul,S.F., Madden,T.L., Schaeffer,A.A., Zhang,J., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*., **25**, 3389–3402.
2. Bancroft,I. (2000) Insights into the structural and functional evolution of plant genomes afforded by the nucleotides of chromosomes 2 and 4 of *Arabidopsis thaliana*. *Yeast*, **17**, 1–5.
3. Eichler,E.E. and Sankoff,D. (2003) Structural dynamics of eukaryotic chromosome evolution. *Science*, **301**, 793–797.
4. Kent,W.J., Baertsch,R., Hinrichs,A., Miller,W. and Haussler,D. (2003) Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. *Proc. Natl Acad. Sci. USA*, **100**, 11484–11489.
5. Goureau,A., Garrigues,A., Tosser-Klopp,G., Lahbib-Mansais,Y., Chardon,P. and Yerle,M. (2001) Conserved synteny and gene order difference between human chromosome 12 and pig chromosome 5. *Cytogenet. Cell Genet*., **94**, 49–54.
6. Lopez-Corrales,N.L., Sonstegard,T.S. and Smith,T.P. (1998) Comparative gene mapping: cytogenetic localization of PROC, EN1, ALPI, TNP1, and IL1B in cattle and sheep reveals a conserved rearrangement relative to the human genome. *Cytogenet. Cell Genet*., **83**, 35–38.
7. Martins-Wess,F., Voss-Nemitz,R., Drogemuller,C., Brenig,B. and Leeb,T. (2002) Construction of a 1.2-Mb BAC/PAC contig of the porcine gene RYR1 region on SSC 6q1.2 and comparative analysis with HSA 19q13.13. *Genomics*, **80**, 416–422.
8. Elgar,G., Sandford,R., Aparicio,S., Macrae,A., Venkatesh,B. and Brenner,S. (1996) Small is beautiful: comparative genomics with the pufferfish (*Fugu rubripes*). *T.I.G*., **12**, 145–150.
9. McLysaght,A., Enright,A.J., Skrabanek,L. and Wolfe,K.H. (2000) Estimation of synteny conservation and genome compaction between puferfish (Fugu) and human. *Yeast*, **17**, 22–36.
10. Thomas,J.W., Touchman,J.W., Blakesley,R.W., Bouffard,G.G., Beckstrom-Sternberg,S.M., Margulies,E.H., Blanchette,M., Siepel,A.C., Thomas,P.J., McDowell,J.C. *et al*. (2003) Comparative analyses of multi-species sequences from targeted genomic regions. *Nature*, **424**, 788–793.
11. Feuillet,C. and Keller,B. (1999) High gene density is conserved at syntenic loci of small and large grass genomes. *Proc. Natl Acad. Sci., USA*, **96**, 8265–8270.
12. Bennetzen,J.L. and Ramakrishna,W. (2002) Numerous small rearrangements of gene content, order and orientation differentiate grass genomes. *Plant Mol. Biol*., **48**, 821–827.
13. Song,R., Llaca,V. and Messing,J. (2002) Mosaic organisation of orthologous sequences in grass genomes. *Genome Res*., **12**, 1549–1555.
14. Acarkan,A., Roßberg,M., Koch,M. and Schmidt,R. (2000) Comparative genome analysis reveals extensive conservation of genome organisation for *Arabidopsis thaliana* and *Capsella rubella*. *Plant J*., **23**, 55–62.
15. O'Neill,C.M. and Bancroft,I. (2000) Comparative physical mapping of segments of genome of *Brassica oleracea* var. *alboglabra* that are homoeologous to sequenced regions of chromosomes 4 and 5 of *Arabidopsis thaliana*. *Plant J*., **23**, 233–243.
16. Parkin,I.A.P., Lydiate,D.J. and Trick,M. (2002) Assessing the level of colinearity between *Arabidopsis thaliana* and *Brassica napus* for *A.thaliana* chromosome 5. *Genome*, **45**, 356–366.
17. Grant,D., Cregan,P. and Shoemaker,C. (2000) Genome organization in dicots: genome duplication in *Arabidopsis* and synteny between soybean and *Arabidopsis*. *Proc. Natl Acad. Sci., USA*, **97**, 4168–4173.
18. Foster-Hartnett,D., Mudge,J, Larsen,D., Danesh,D., Yan,H., Denny,R., Peñuela,S. and Young,N.D. (2002) Comparative genomic analysis of sequences sampled from a small region on soybean (*Glycine max*) molecular linkage group G. *Genome*, **45**, 634–645.
19. Yan,H.H., Mudge,J., Kim,D.J., Larsen,D., Schoemaker,R.C., Cook,D.R. and Young,N.D. (2003) Estimates of conserved microsynteny among the genomes of *Glycine max*, *Medicago truncatula* and *Arabidopsis thaliana*. *Theor. Appl. Genet*., **106**, 1256–1265.
20. Mayer,K., Murphy,G., Tarchini,R., Wambutt,R., Volckaert,G., Pohl,T., Düsterhöft,A., Stiekema,W., Entian,K.-D., Terryn,N. *et al*. (2001) Conservation of microstructure between a sequenced region of the genome of rice and multiple segments of the genome of *Arabidopsis thaliana*. *Genome Res*., **11**, 1167–1174.
21. Salse,J., Piégu,B., Cooke,R. and Delseny,M. (2002) Synteny between *Arabidopsis thaliana* and rice at the genome level: a tool to identify conservation in the ongoing rice genome sequencing project. *Nucleic Acids Res*., **30**, 2316–2328.
22. Ku,H.-M., Vision,T., Liu,J. and Tanksley,S.D. (2000) Comparing sequenced segements of tomato and *Arabidopsis* genomes: large-scale duplication followed by selective gene loss creates a network of synteny. *Proc. Natl Acad. Sci., USA*, **97**, 9121–9126.
23. Gebhardt,C., Walkemeier,B., Henselewski,H., Barakat,A., Delseny,M. and Stüber,K. (2003) Comparative mapping between potato (*Solanum tuberosum*) and *Arabidopsis thaliana* reveals structurally conserved domains and ancient duplications in the potato genome. *Plant J*., **34**, 529–541.
24. Dunham,I., Shimizu,N., Roe,B.A., Chissoe,S., Hunt,A.R., Collins,J.E., Bruskiewick,R., Beare,D.M., Clamp,M., Smink,L.I. *et al*. (1999) The DNA sequence of human chromosome 22. *Nature*, **402**, 489–496.
25. Chromosome 21 mapping and sequencing consortium (2000) The DNA sequence of human chromosome 21. *Nature*, **405**, 311–319.
26. International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
27. Venter,J.C., Adams,M.D., Myers,E.W., Li,P.W., Mural,R.J., Sutton,G.G., Smith,H.O., Yandell,M., Evans,C.A., Holt,R.A. *et al*. (2001) The sequence of the human genome. *Science*, **291**, 1304–1351.
28. Mungali,A.J., Palmer,S.A., Sims,S.K., Edwards,C.A., Ashurst,J.L., Wilming,L., Jones,M.C., Horton,R., Hunt,S.E., Scott,C.E. *et al*. (2003) The DNA sequence and analysis of human chromosome 6. *Nature*, **425**, 805–812.
29. Lin,X., Kaul,S., Rounsley,S., Shea,T.P., Benito,M.-I., Town,C.D., Fujii,C.Y., Mason,T., Bowman,C.L., Barnstead,M. *et al*. (1999) Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. *Nature*, **402**, 761–768.
30. Mayer,K., Schüller,C., Wambutt,R., Murphy,G., Volckaert,G., Pohl,T., Düsterhöft,A., Stiekema,W., Entian,K.-D., Terryn,N. *et al*. (1999) Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*. *Nature*, **402**, 769–777.
31. Salanoubat,M., Lemcke,K., Rieger,M., Ansorge,W., Unseld,M., Fartmann,B., Valle,G., Blöcker,H., Perez-Alonso,M., Obermaier,B. *et al*. (2000) Sequence and analysis of chromosome 3 of the plant *Arabidopsis thaliana*. *Nature*, **408**, 820–822.
32. Tabata,S., Kaneko,T., Nakamura,Y., Kotani,T., Kato,T., Asamizu,E., Miyajima,N., Sasamoto,S., Kimura,T., Hosouchi,T. *et al*. (2000) Sequence and analysis of chromosome 5 of the plant *Arabidopsis thaliana*. *Nature*, **408**, 823–826.
33. Theologis,A., Ecker,J.R., Palm,C.J., Federspiel,N.A., Kaul,S., White,O., Alonso,S.Y., Altafi,H., Araujo,R., Bowmann,C.L. *et al*. (2000) Sequence and analysis of chromosome 1 of the plant *Arabidopsis thaliana*. *Nature*, **408**, 816–819.
34. Chain,P., Kurtz,S., Ohlebush,E. and Slevak,T. (2003) An applications-focused review of comparative genomics tools: capabilities, limitations, and future challenges. *Brief. Bioinformatics*, **4**, 1–20.
35. Mullikin,J.C. and Ning,Z. (2003) The phusion assembler. *Genome Res*., **13**, 81–90.
36. Couronne,O., Poliakov,A., Bray,N., Ishkhanov,T., Ryaboy,D., Rubin,E., Pachter,L. and Dubchak,I. (2003) Startegies and tools for whole-genome alignments. *Genome Res*., **13**, 73–80.
37. Birney,E., Andrews,D., Bevan,P., Caccamo,M., Cameron,G., Chen,Y., Clarke,L., Coales,G., Cox,T., Cuff,J. *et al*. (2004) Ensembl 2004. *Nucleic Acids Res*., **32**, D468–D470.
38. Wheeler,D.L., Church,D.M., Federhen,S., Lash,A.E., Madden,T.L., Pontius,J.U., Schuler,G.D., Schriml,L.M., Sequeira,E., Tatuva,T.A. *et al*. (2003) Database resources of the National Center for Biotechnology. *Nucleic Acids Res*., **31**, 28–33.
39. Schoof,H., Ernst,R., Nazarov,V., Pfeifer,L., Mewes,H.W. and Mayer,K.F. (2004) MIPS *Arabidopsis thaliana* Database (MatDB): an integrated biological knowledge resource for plant genomics. *Nucleic Acids Res*., **32**, D373–D376.

40. Rice,P., Longden,I. and Bleeasby,A. (2000) EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet*., **16**, 276–277.
41. Betts,M.J., Guigo,R., Agarwal,P. and Russell,R.B. (2001) Exon structure conservation despite low sequence similarity: a relic of dramatic events in evolution? *EMBO J*., **20**, 5354–5360.
42. Rozen,S. and Skaletsky,H.J. (2000) Primer3 on WWW for general users and for biologist programmers. In Krawetz,S., Misemer,S. (ed.), *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowa, NJ, pp. 365–386.
43. Demeure,O., Renard,C., Yerle,M., Faraut,T., Riquet,J., Robic,A., Schiex,T., Rink,A. and Milan,D. (2003) Rearranged gene order between pig and human in a QTL region on SSC 7. *Mamm. Genome*, **14**, 71–80.
44. Robic,A., Faraut,T., Iannuccelli,N., Lahbib-Mansais,Y., Cantegrel,V., Alexander,L. and Milan,D. (2003) A new contribution to the integration of human and porcine genome maps: 623 new points of homology. *Cytogenet. Genome Res*., **102**, 100–108.
45. Bosak,N., Faraut,T., Mikawa,S., Uenishi,H., Kiuchi,S., Hiraiwa,H., Hayashi,T. and Yasue,H. (2003) Construction of a high-resolution comparative gene map between swine chromosome region 6q11–>q21 and human chromosome 19 q-arm by RH mapping of 51 genes. *Cytogenet. Genome Res*., **102**, 109–115.
46. Morisson,M., Jiguet-Jiglaire,C., Lemiere,A., Leroux,S., Faraut,T., Yerle,M. and Vignal,A. (2003) A radiation hybrid panel and its use in developing a gene map of the chicken. *Br. Poult. Sci*., **44**, 797–798.
47. Hayes,H., Elduque,C., Gautier,M., Schibler,L., Cribiu,E. and Eggen,A. (2003) Mapping of 195 genes in cattle and updated comparative map with man, mouse, rat and pig. *Cytogenet. Genome Res*., **102**, 16–24.
48. Li,L., Stoeckert,C.J.,Jr and Roos,D.S. (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res*., **13**, 2178–2189.
49. Cannon,S.B. and Young,N.D. (2003) OrthoParaMap: distinguishing orthologs from paralogs by integrating comparative genome data and gene phylogenies. *BMC Bioinformatics*, **4**, 35.
50. Leveugle,M., Prat,K., Perrier,N., Birnbaum,D. and Coulier,F. (2003) ParaDB: a tool for paralogy mapping in vertebrate genomes. *Nucleic Acids Res*., **31**, 63–67.
51. Wolfe,K.H. (2001) Yesterday's polyploids and the mystery of diploidization. *Nature Rev*., **2**, 333–341.