QTL discovery for yield, disease and pest traits in Populus. Chapter 3

Véronique Jorge, Véronique Storme, Anne M Rae, Muriel Gaudet, Isabella Paolucci, Catherine Bastien, Gail Taylor, Maurizio Sabatti

DETAILED REPORT OF THE PROJECT POPYOMICS
CONTENTS

EXECUTIVE SUMMARY

Contract:QLK5 –CT- 2002-00953 Duration: **Nov 1st, 2002 – May 31, 2006**

Title: **Linking physiology, molecular genetics and genomics in *Populus* to understand and improve yield and quality for biomass and timber production across Europe (POPYOMICS)**

Objectives

To develop a new experimental network for Europe, consisting of three sites: a) in southern UK, b) in central France and c) in northern Italy. Each will consist of four 'mapping pedigrees' (POP1-4) of *Populus* planted to rigorous experimental design. To identify QTL in four mapping pedigrees (POP1, POP2, POP3, POP4) for growth and vigour, disease resistance and timber yield using available molecular genetic maps and available software. To test the robustness of QTL in four mapping pedigrees (POP1, POP2, POP3, POP4) by growth of pedigrees at the sites with contrasting climates, thus examining genotype x environment interactions.. In an innovative approach, to use molecular markers to develop a POPLAR consensus map. To achieve this by the use of available SSR markers in the public domain, and to validate their usefulness in all four pedigrees. To use association (linkage disequilibrium mapping) to test the co-location of markers and QTL using two collections of the native *P. nigra* available in the UK, Belgium and Italy. To utilise the DNA microarrays available for wood and leaf ESTs in *Populus* with the parents of the poplar pedigrees, with focus on POP1, to identify differentially expressed ESTs in response to abiotic and biotic stresses and to map these. To use the EST database to identify up to 100 candidate genes for traits associated with yield and diseases resistance and to map these candidate genes in the appropriate *Populus* pedigree. To confirm the linkage group locations of these candidate genes from full-length DNA sequences, as available, from the emergent *Populus* physically mapping project on-going in the USA.

Scientific achievements

We have identified a set of traits that are able to effectively describe the biomass yield of a wide range of *Populus* genotypes. These traits were measured in several thousand trees at three sites including a large destructive harvest that enables us to link traits to yield, following three years of sampling. This close attention to detail has been followed through by the development of a PHYSIO-TRAIT database that is currently being prepared for access by the public and will provide a useful resource. A similar database is in development for disease traits, particularly those related to rust. Both databases will be part of the virtual laboratory, 'The laboratory without walls for ecosystem genomics'. In establishing the largest natural collection of *Populus nigra*, we were able to assess nucleotide diversity in this species for the first time and utilise a candidate gene approach to undertake association analysis. The 500 genotypes of *P. nigra* were planted in a fully replicated design and enable us to follow bud-set over several weeks during the 2006 autumn. These data are unique and have enables us to link this developmental trait to underlying genes for the first time in such a tree population. Several genes, particularly those related to the switch to flowering in Arabidopsis, appear to have a role in this process.

QTL discovery has identified one area of the genome from six different maps, that is involved in bud burst and similarly QTL for yield and syllepsis. These QTL explain up to 17. 2 % of the variance associated for a trait. QTL for rust for both quantitative and qualitative resistance have also been identified accounting for a considerable portion of trait variation. These data and those for other QTL are available for display using the program cMAP located on a server

that will become publicly available after publication of the data within the project. They are the first large-scale cross-pedigree study to be available in this way.

Genetical genomic approaches were used extensively in POPYOMICS. Two partners contributed to the *Populus* international sequencing effort during the course of the project and two partners were also involved in developing *Populus* cDNA microarrays which were used to link gene expression to geentics by following expression in progeny from the mapping population. In response to drought a set of genes was identified that were differentially expressed between the paprents of mapping progeny and these have been used as candidates in a large-scale RT-PCR (real time PCR study), enabling expressionQTL (eQTL) to be mapped. This is a powerful approach which can help us to elucidate *cis* and *trans* acting factors in the regulation of gene expression in realtion to traits of interest and this approach is also being developed to understand the gene underlying rust tolerance in *Populus*.

Main deliverables

POPYOMICS was unique in bringing together expertise in a timely way. The experimental network created at three sites was world-leading and has never before been achieved, since the experiment spanned three contrasting climates in Europe and where we planted five *Populus* pedigrees in an identical replicated design. The dataset from these plantations has revealed important G x E interactions for traits and identified a set of QTL for *Populus* yield that are robust and of direct value for future breeding and commercialisation. Simialrly, QTL for rust resistance were identified and those QTL sensitive to different environments have also been found. Coupled with this, in BASE2 and within the French transcriptome database, we have identified a suite of *Populus* candidate genes for response to drought, rust infection, ozone, and a variety of other environmental challenges and have also linked these gene expression patterns to QTL. For the first time in *Populus* and at a scale of 500 genotypes, we used association genetics to link important phenology triats, in this instance, bud-set, to underlying genes and have identified clearly genes likely to have a role in this developmental process.

Socio-economic relevance and policy implications

Europe has set ambition policy targets for renewable energy for heat, power and liquid transportation fuels, suggesting that 30-40 % of primary supply may be met from this source by 2020. For liquid transportation alone a target of 5.75 % rising to 10% is a requirement of the RTFO renewable transport fuel obligation by 2010 and 2020 respectively. Biomass is unique in being able to provide a source of feedstock for all of these energy conversions, delivering energy that is sustainable and close to carbon neutrality. Of the renewable technologies identified, biomass has immediate potential and second generation lingo-cellulosic crops such as *Populus*, grown exclusively for energy and not food requirement, are likely to dominate the landscape in marginal areas of Europe over the coming decades with important consequences for the rural environment, employment and in helping to meet the EU policy targets. It is well recognized that a bottle neck currently exists with such crops in that yield is too low and that in future, these largely undomesticated trees must be improved rapidly through non GM biotechnological routes. Within this framework, the outputs of POPYOMICS will have long-term durability and help to add momentum to this important area of benefit to Europe in future.

Conclusions

The project POPYOMICs provided a timely opportunity to undertake a large integrated study on poplar bringing together breeders, physiologists, molecular biologists and geneticists. This was because during the lifetime of the project, the first tree, *Populus* was fully sequenced at the level of DNA, a project involving partners from POPYOMICS working in collaboration as part of a large USA DOE project. As a consequence POPYOMICS benefitted from early access to sequence information and was able to make full use of this in linking QTL data to that of the DNA sequence and providing a platform to utilize genetical genomics to link traits to genes, which has rarely been done in trees prior to this project.

Our outputs include a new comparative map across five *Populus* pedigrees and this is of wide value in identifying traits for future breeding that are linked to yield and disease resistance. Linked to this we made important advances in the genetical genomics approach, using the recently available poplar microarrays and employing them widely across the project for gene discovery. Association genetic is a powerful approach to like traits to genes and here we were able to develop 'proof' of concept that this is a valuable technology for associating phenology traits to underlying genes. All of these resources including a trait databases. QTL databases and gene expression databases are being developed further in three new research projects including the FP6 NOE, 'EVOLTREE', the FP7 projects 'ENERGY POPLAR' and 'NOVELTREE' and enabling us to main international links with the USA and Canada. These project will enable the valuable outputs of POPYOMICS to be captured and exploited for future low carbon biofuels and in ensuring the conservation of Europe's forest resource in the face of climate change.

Keywords: *Populus*, yield, bioenergy, pest and disease resistance, QTL, genetics, genomics

1. Background, scientific objectives and development of a European experimental network for *Populus*

M. Sabatti, C. Bastien, R. Ceulemans, AM Rae, NR Street, M Pinel, M. Steenackers, L. Ricciotti, G. Scarascia Mugnozza, G. Taylor

1.1 Introduction

Trees for biomass energy and wood in Europe – developing a European Experimental Network

The potential of trees for contribution to the EU commitment to carbon dioxide emissions reduction and for improved energy security is considerable, as identified recently in the European EPOBIO project and associated publications [1] and in the EU roadmap for future development of liquid fuels [2]. This is because these trees not only remove carbon from the environment, but they also displace additional fossil fuel consumption for energy generation. It has been suggested that this displacement effect may account for up to 76 % of the carbon benefit of the trees. It was argued recently that of the proposed land-use changes for carbon mitigation across Europe, woody biomass energy crops offer by far the greatest potential mitigation, again because they both sequester carbon and displace fossil-fuel carbon [3]. In comparison to fossil-fuel based systems that emit approximately 1,200 g CO_2 per kWh energy, energy crops emit, from a net life-cycle analysis, 65 g CO_2 per kWh power.

As such, the POPYOMICS project was extremely timely, providing underpinning data to address the issue of trees as a source of bioenergy, including liquid transportation fuel. One of the current bottle necks for development of these woody crops for bioenergy, is the consistent and reliable production of superior yield [4]. Since these trees are largely undomesticated, very limited improvement has been undertaken and here, the high through-put systems approaches of POPYOMICS have been extremely beneficial. By gathering together the expertise in field biology, physiology, genetics and genomics, POPYOMICS has provided Europe with a leading position in poplar energy bioscience.

An important aspect of the project was the utilisation of *Populus* as both a 'model' and a 'commercially relevant' agroforestry system and our benefit from the availability of the full poplar DNA sequence that was published during the lifetime of the POPYOMICS project [5]

enabling us to make a link between molecular genetic and physical DNA sequence. A central and important part of the POPYOMICS project was to bring together all of the mapping populations in *Populus* from across Europe and beyond and to develop a unique resources in mapping populations for *P. alba* and *P. nigra.* An important part of bringing these genetic resources together was the establishment of an experimental network across Europe and associated plantations that may be summarized as:

- **Experimental Network – four mapping populations of *Populus***

Identical trials were established in three countries (UK, F, I) using five different poplar pedigrees enabling genotype-environment interactions to be investigated to test the robustness of QTL. The sites were planted to the same experimental protocol and design and spanned a wide range of climatic and environmental conditions.

- **Nursery trials of *Populus* for disease resistance**

In addition to the main experimental network, intensive nursery trials in two countries (B, F) for resistance disease traiting were also initiated. These trials are not described in detail here and are dealt with in Chapter 3.

- **The *P. nigra* linkage disequilibrium and association mapping collection**

Finally a unique resource was planted in Belgium – a large collection of *P. nigra* from across Europe, totaling 500 genotypes selected from a diverse range of latitudinal origin.

- **The POPFACE site for allied genomics research**

Measurements were also made at the POPFACE elevated CO_2 plantation experiment in WP5.

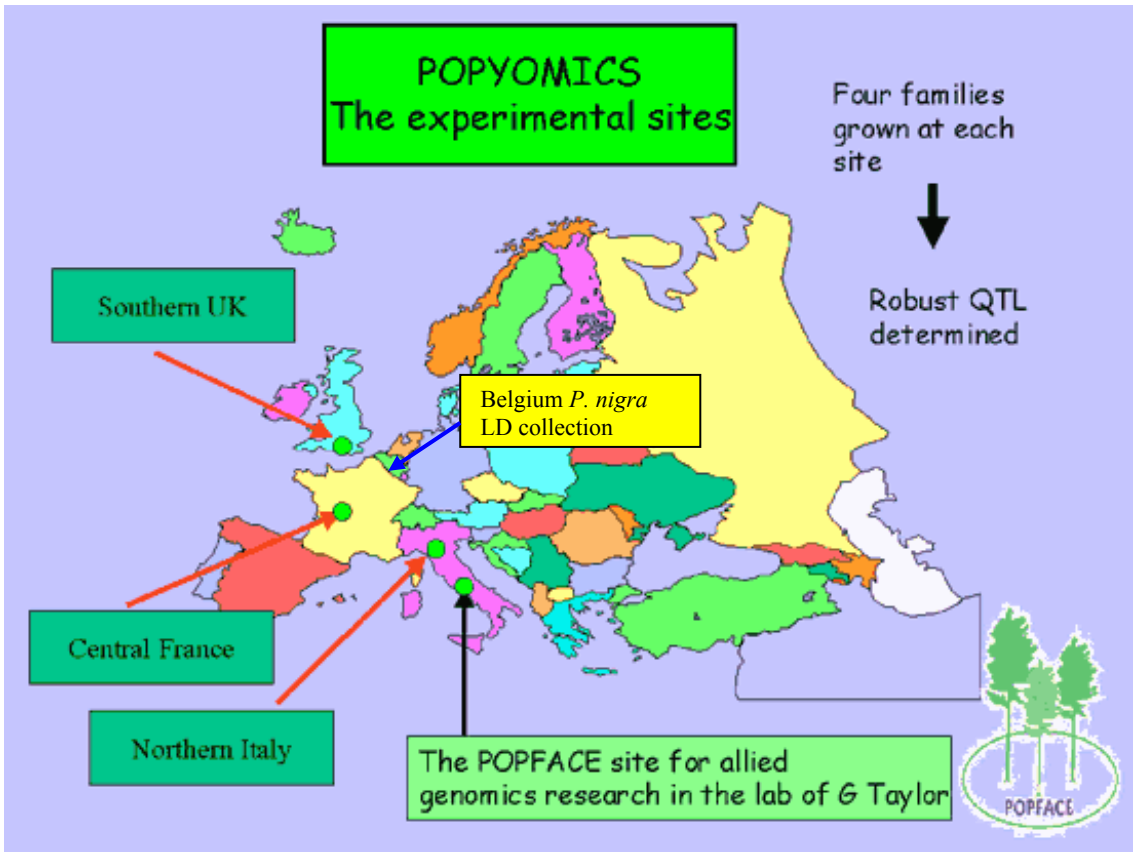


Plate 1.1 The four experimental network sites are shown in the UK, France and Italy. In addition, measurements were made at the POPFACE elevated CO_2 plantation experiment in WP5 and the *Populus nigra* collection was grown in Belgium.

1.2 Site description, planting and management

Experimental Network – four mapping populations of Populus

The genetic resource available for this project was extensive, some of which has been reported elsewhere [6] and some of which was developed within POPYOMICS, for example the mapping pedigrees of POP4 and POP5 by the Italian partners in Viterbo. Wide genetic background combined with plantations established in contrasting climatic zones enabled the POPYOMICS project to investigate genotype x environment (G x E) interactions for important biomass and disease and pest resistance traits [7]. The material available to the project is summarised in Table 1.1 For the first time, here different pedigrees of poplar were brought together, within one project and experimental trials coordinated for optimization of data collection. In general, traditional breeding in trees suffers from the extremely long generation time, slow growth and difficulty in the selection of superior parents. It must be taken into account that these materials

are the result of routine but important activities, such as breeding and maintenance of living clone banks, difficult to justify to most funding agencies. For this reason, the materials available for this project represented an important resource and a considerable scientific investment for the future.

The pedigrees consisted of F_1 and F_2 generations *ad hoc* obtained selecting parental genotypes to identify polygenic yield traits and poly- and monogenic traits for disease resistance. The original parental material belonging to two North American, *P. deltoides* (D) and *P. trichocarpa* (T), and two native species from Europe, *P. nigra* (N) and *P. alba* (A). One native pure (AxA) and three hybrid progenies (TxD, DxT and DxN) involving more than 1000 genotypes were used to establish the three experimental plantations in United Kingdom, France and Italy in an oceanic, continental and Mediterranean climate, respectively.

Table 1.1 A summary of the genetic material utilised for the experimental network in POPYOMICS. POP4 and POP5 were developed within the project.

Pedigree code	Partner responsible	Original pedigree code	Parentage	Progeny number / generation	Origin at start of project
POP1	01	331	*P.trichocarpa* x *P. deltoides*	210 - F_2	UK
POP2	03	54B	*P.deltoides x P.trichocarpa*	350 - F_1	France
POP3a POP3b	05	87.001 87.002	*P.deltoides x P.nigra* *P.deltoides x P.trichocarpa*	225 - F_1	Belgium and UK
POP4	06	14P6K	*P.alba* x *P.alba*	150 - F_1	Italy
POP5	06	NA	*P. nigra* x *P. nigra*	180 – F_1	Italy

Work on many physiological and morphological traits linked to yield and susceptibility to biotic constraints was performed using the available pedigrees, as described in chapters 2, 3 and 4 of this report and detailed briefly in Table 1.2.

Table 1.2 Measurements made prior to POPYOMICS and map status available for the five poplar pedigrees.

Pedigree code	Partner	Traits studied in prior work	Traits to measure for this project	Map status
POP1	01	Traits of physiology and morphology linked to yield but only in USA. Rust resistance QTL identified (mono-morphic to *Melampsora. meduseae)*.	Several QTL for physiological and morphological traits linked to stem wood yield have already been identified in this population. The objective here will be to test physiological and morphological traits for QTL robustness when the pedigree is grown in a wide range of European climate conditions.	Over 700 markers published. Most importantly SSR markers available from USA lab to this project, allowing alignment of poplar maps.
POP2	04	Resistance for rust and basic growth traits measured.	Replicated experiment with and without chemical control of the rust pathogen *Melampsora spp*. will be performed using this material in three experimental sites, allowing the rust resistance QTL to be identified.	Map in progress, 360 markers for D and 290 for T. Integration of SSR marker available from the project
POP3	05	Resistance for rust mapped. Basic growth traits measured.	The purpose of the experimental trials will be to define more closely disease resistance and leaf and branch traits for yield on material of the whole pedigree.	Map published, several hundred AFLP and SSR markers.
POP4	06	Traits of physiology and morphology linked to yield measured.	The parents, selected from the north and south of Italy, are strikingly different for phenology and crown traits important determinants of the yield that will be studied in different site conditions to test QTL robustness.	Map in progress, selected SSR, RFLP and RAPD markers. Integration of SSR marker available from the project.
POP5	06	None	Leaf development traits and yield traits	Map published, developed entirely within POPYOMICS

Three very large experimental trials in I, F and UK were established with in excess of 30,000 poplar cuttings generated by the partners, which were distributed and planted over the winter 2002 – 2003. The establishment of the clonal material at the sites was excellent in the UK and F and less stable in I. In particular, the establishment of POP1 and POP3a was less good at the southern Italian site, likely to reflect the small size of the pencils and the limited irrigation system in place at that site, coupled with the fact that one of the parents of the POP1 is selected from an extremely wet river corridor in the western USA [4]. Nevertheless, useful traiting has been completed at this site for the POP1 and POP3a pedigrees. These sites provided an outstanding and unique resource for Europe both within and external to POPYOMICS. The POPYOMICS

project was open to other scientists and a number of additional data sets were collected from the experimental network during the POPYOMICS project.

Throughout the four years of the project, material was collected, distributed, planted and managed by three of the nine partners. The tasks undertaken included: *Site preparation*. A broad-spectrum contact herbicide was used prior to planting to control weeds at each site. *Planting.* A uniform design protocol was followed as detailed below, with hard-wood pencils inserted directly into the ground after pre-soaking in water for 24 h. *Cut-back.* At the end of YEAR 2, the trees/plants were cut-back to allow a coppice to develop during YEAR 3. This facilitated traiting for physiological, phenotypic, growth and disease traits. *Protection.* The experiments were protected from rabbit and deer damage by suitable fencing. *Post-planting weed control.* This was necessary only during the first year of establishment, before canopy closure and was achieved with a contact herbicide.

P1, P3 and P6 agreed, during the course of POPYOMICS that they would assume responsibility for UK (POP1), F (POP2) and I (POP4) respectively for the Network of Excellence, EVOLTREE (www.evoltree.org), that kicked-off in April 2006 as part of Framework 6, since when limited funding has been available to maintain the trees in stool beds but not to support the experimental network, which at the end of the project was removed.

Table 1.3 Diary of events for the establishment and maintenance of the POPYOMICS experimental network.

Date	Activity
November 2002	Up to 30,000 cuttings produced in the UK, F, I and B.
January 2003	Cuttings organised and swapped between partners
January 2003	Final experimental planting design agreed
March 2003	Pre-emergent weedkiller
March-April 2003	Plantations established
June 2003	Singled stems
January 2004	Cut-back and re-growth of single stem
January 2006	Final harvest
Throughout	Irrigation and effective weed and rust management

The Experimental Design

It was agreed that a common experimental design and protocol should be used for the network that would enable analysis and synthesis of data and the development of joint publications. The

initial proposition of experimental design was discussed in December 2002 during the first meeting in London (UK) and re-adjusted with exact number of genotypes in January 2003.

There were 2 types of experiments. **Type 1** was composed of 12 complete blocks, 6 blocks treated with fungicide and 6 blocks untreated to jointly evaluate growth potential, resistance and tolerance to rust infection. **Type 2** was composed of 6 (9 for *P. alba* pedigree, POP4) complete blocks treated with fungicide during all the project to avoid rust infection. Type 1 was only employed in the two sites (F, I) for POP3a and POP3b pedigrees (sub-populations of POP3, reflecting the two crosses of the F_1 pseudo testcross) enabling genotype-environment interactions for rust resistance to be investigated. Type 2 was used for the three sites of the three other pedigrees POP1, POP2 and POP4.

For a given pedigree, 6 **'bridge genotypes**' were defined. They were used as standard clones in all trials and to make bridges between trials. The parents were tested only in the 3 sites devoted to the corresponding pedigree. Reference clones covering a range of rust susceptibility were included in type 1 experiments.

Plate 1.2 An overview of activity for gathering of genotypes, preparation of cuttings, planting cuttings and first year growth during the experimental network.

All experiments were protected from rust infection during the first growing season (2003) and were irrigated according to needs. Fungicide was applied on type 2 experiments and on the 6 treated blocks of type 1 experiments during the second growing season (2004) in order to evaluate ecophysiological components and growth potential in absence of rust infection. All experiments were coppiced during winter 2004-2005. Initial spacing was 2m between rows, 0.75m between plants on a row. Two border rows were planted around each trial and 5 border rows were included between treated and non treated blocks in type 1 experiments.

Developing a new collection of P. nigra for linkage disequilibrium analysis and association mapping.

In order to perform a linkage disequilibrium (LD) association mapping of quantitative traits, a *P. nigra* mapping population was established in Belgium. The constituents of the wide population originated from the EUROPOP (EU FP5-project) and EUFORGEN *P. nigra* collection and the private collection of partner 3. Populations were selected along different riverine systems: the Rhine (populations from The Netherlands and Germany), the Loire, Drôme and Durance (populations from France), the Ticino (the Italian population) and the Ebro (the Spanish population).

Cuttings from all genotypes were kindly provided by the following researchers:

- Dr. Stefano Castiglione, Università degli Studi di Milano, Dipartimento di Biologia, Milano, **ITALY**
- Dr. Carmen Maestro, Centro de Investigación y Tecnología Agroalimentaria (CITA) , Zaragoza, **SPAIN**
- Dr. Catherine Bastien, INRA, Orléans, **FRANCE** (Partner 3)
- Mr. Sven De Vries, Alterra, Wageningen, **THE NETHERLANDS**
- Dr. Karl Gebhardt, Hessen-Forst, Hann. Münden, **GERMANY**

Table 1.4 Main *P.nigra* populations

Country	**Population**	**Nbr. of Clones**	**Riverine system**
Spain	POP1	31	Ebro
	POP2	44	Ebro
Italy	Bosco Siro Negri	44	Ticino
	La Zelata	63	Ticino
France	POP1	31A + 32S	Drôme
	POP6	32A + 31S	Drôme
Germany	POP1	56	Rhine
The Netherlands	POP1	50	Rhine

Table 1.5. Minor *P.nigra* populations

Nbr. of Clones	Riverine system	Population
12	Durance	Various
26	Loire (West of Orléans)	Various
21	Loire (West of Orléans)	Various
8 standard clones*		

****** French clones representative for *P.nigra* variation for rust resistance and growth potential.

For each population, six cuttings per genotype were provided by Spain, Italy and France in January 2004. The *P.nigra* populations from Germany and the Netherlands were already present in the experimental nursery of partner 5 since 2002. For these populations, cuttings were collected from mother-stocks produced in the nursery of partner 5. All cuttings were stored in a cold chamber at +2°C until planting. All cuttings were labelled individually and packed by block according to the experimental design **Type 2** as adopted for the network of field trials established in I, UK and F in 2003: six complete randomised blocks treated with fungicide for the duration of the project to prevent rust infection. Supplemental standard clones, representative for *P.nigra* variation for rust resistance and growth potential, were provided by partner 3 and 6 to be included in the experiment.

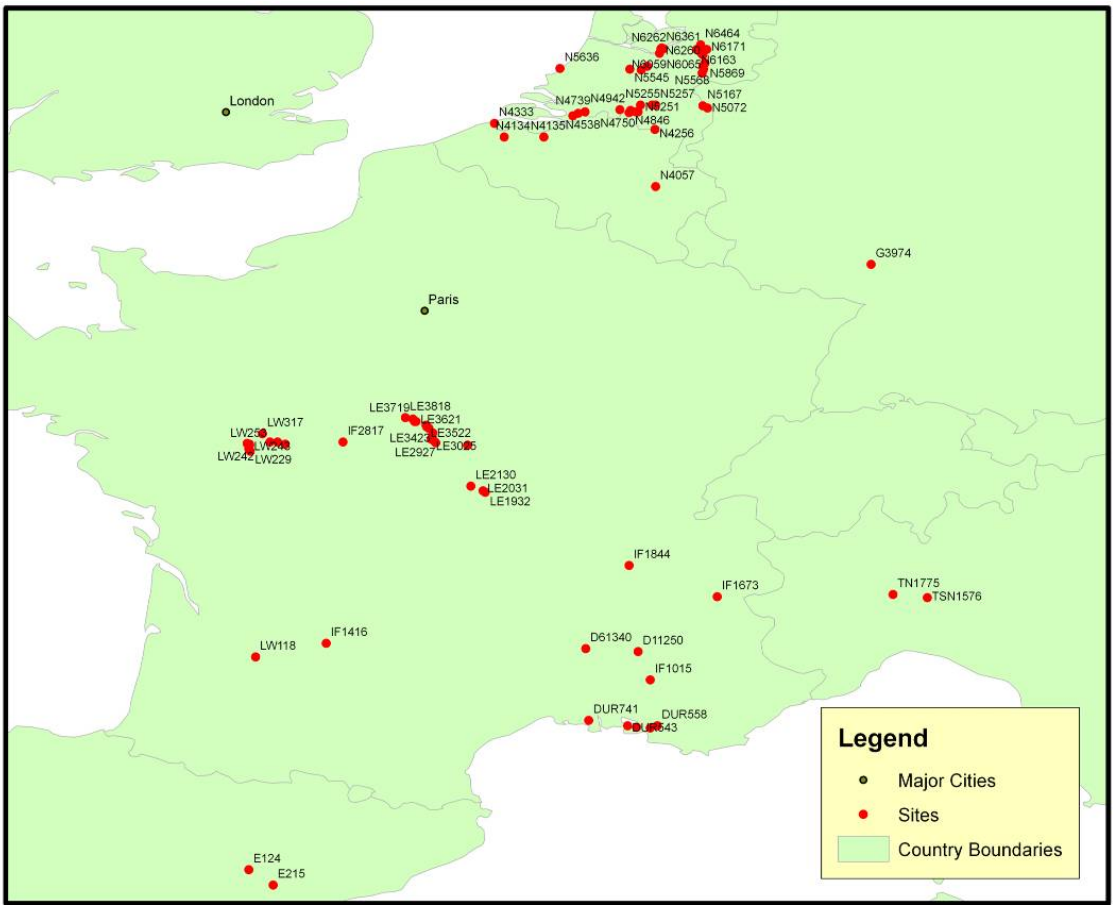


Plate 1.3 A map to show the places of origin of the *P. nigra* collection. Each red spot indicates a collection site, with codes for each population given.

Establishment of the field trial using the different P. nigra populations

The experimental plantation in Belgium was established in April 2004. In March, the soil was ploughed and prepared for planting. A broad range contact herbicide was applied prior to planting to control weeds. Hardwood cuttings were pre-soaked in water for 48h before planting. Planting was done according to the experimental design. For 56 clones, cuttings were too small to be planted immediately in the field trial. These cuttings were planted in nursery trays and were transplanted to the field test in spring 2005. A plan of the experiment was drawn and verified. All plants were labelled individually. At the beginning of July 2004, the number of missing or dead plants was recorded. During the vegetation period manual weed control appeared to be necessary and was performed twice. Rust was treated by spraying every three weeks from mid July till September using Horizon (Tebuconazole).

Results

Good survival rates and satisfactory growth were observed in the experimental trial (Table 1.6). In September, 2-3 young leaves were collected from all genotypes in one block. Leaves were wrapped up in foil and stored immediately first in liquid nitrogen in the field, later at -80°C in the lab. Subsequently, leaves were lyophilized and sent to partner 9 for DNA extraction.

Table 1.6 Survival rate in the LD experimental site at Geraardsbergen (Belgium)

COUNTRY	CLONES	Plants	Survival Rate %
Spain	75	450	87.1
The Netherlands	50	300	98.3
Germany	56	333	89.8
Italy	107	642	92.4
France	192	1152	93.5
POP5	8	48	97.9
TOTAL	**488**	**3567**	**92.4**

Organising this collection was more difficult than anticipated since the trees are dispersed widely across Europe and in the hands of many private as well as institutional growers. Despite this, progress was finally made, but not in time to allow planting of this trial in year 1 of the project – this is because planting could only occur during the months of February to April. Given the difficulties of obtaining adequate material to make a single LD trial, this collection was initially only planted in B and not additionally in the UK. The science from the study was still appropriate although efforts are still being made to plant this trial in the UK, for future analysis.

1.4 A summary of the main uses of the experimental trials and plant material

The plant material described in this chapter underpinned all of the research that formed the POPYOMICS project. Mapping populations are a critical component of a novel approach used here which was that of genetical genomics [8]. In genetical genomics, high throughput

transcriptomics are performed on progeny from mapping population, either individuals or as in our case, on extreme genotypes – those showing differences in yield or response to stresses such as drought [9]. Initially this approach was largely confined to human health studies where much can be discovered about the genetic architecture and control of a plant trait. Jansen and Nap described the approach of genetical genomics as 'harnessing the power of genetics to the resources of genomics' [8]. For plants in general and trees in particular, few genetical genomics studies have yet to be reported, although several are underway and in the first large-scale analysis of eQTL [10] it appears that the genetic architecture of plant traits may be complicated both *cis* and *trans* acting effects apparent.

A second major use of the plant material described here was in the discovery of genotype-environment interactions. For the first time, more than one mapping population of *Populus* was brought together at more than one site, providing a timely opportunity to discovery robust and plastic QTL and deduce their relevance for adaptation to climate and in future breeding and improvement programmes.

Third and finally, association genetics provides a powerful alternative to QTL mapping since the population is not limited by recombination events which may be an issue when mapping QTL in trees [11]. The development of the *Populus nigra* collection described here, the largest single natural collection of *Populus* globally, has provided a world class resource from which to deduce associations between adaptive traits, in this instance, bud set to underlying genes of interest [12].

1.5 Conclusions

Five mapping populations for QTL elucidation, planted at three sites to an identical experimental design were established and successfully managed within POPYOMICS. They were used to gain insight into G x E interactions in *Populus* and elucidate robust QTL. A set or nursery trials was developed to establish the genetic basis of disease tolerance and resistance and a unique natural collection of *Populus nigra* was developed, planted and managed in an experimental field trial. In all, in excess of 30,000 trees were utilized throughout the project.

1.6 References

1. Moller, R, Toonen M, Van Beilen J, Salentijn E, Clayton D (2007). Crop Platforms for cell wall biorefining: lignocellulosic feedstocks. EPOBIO, http://www.epobio.net/pdfs/0704LignocelluloseFeedstocksReport.pdf.

2. Biofuels in the European Union: A vision for 2030 and beyond. Final Report of the Biofuels Research Advisory Council, 2006, http://www.refuel.eu/fileadmin/refuel/user/docs/BIOFRACbiofuels_vision_2030_en.pdf

3. Smith P, Powlson DS, Smith JU, Falloon P, Coleman K (2000). Meeting Europe's climate change commitments: quantitative estimates of the carbon mitigation potential of agriculture.

4. Rae AM, Robinson KM, Street N and Taylor G (2004). Structural, morphological and biochemical traits linked to fast growth in short-rotation coppice poplar for biomass Canadian Journal of Forest Research, 34(7): 1488-1498.

5. Tuskan G, DiFazio S, Jansson S. et al. (2006) The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). Science 313: 1596–1604.

6. Cervera MT, Storme V, Ivens B, Gusmao J, Liu BH, Hostyn V, Van Slycken J, Van Montagu M, Boerjan W (2001) Dense genetic linkage maps of three Populus species (Populus deltoides, P-nigra and P-trichocarpa) based on AFLP and microsatellite markers. Genetics, 158, 787-809.

7. Rae AM, Pinel MPC, Bastien C, Sabatti M, Street NR, Tucker J, Dixon C, Marron N, Dillen SY and Taylor G (in press) QTL for yield in bioenergy Populus: identifying GxE interactions from growth at three contrasting sites. Tree Genetics & Genomes.

8. Jansen RC Nap JP (2001) Genetical genomics: the added value from segregation. Trends in Genetics, 17, 388-391.

9. Street NR, Skogstrom O, Sjodin A, Tucker J, Rodriguez-Acosta M, Nilsson P, Jansson S and Taylor G (2006). The genetics and genomics of the drought response in Populus. Plant Journal, 48 (3): 321-341.

10. West MAL, Kim K, Kliebenstein DJ, van Leeuwen H, Michelmore RW, Doerge RW and Clair DAS (2007). Global eQTL mapping reveals the complex genetic architecture of transcript-level variation in Arabidopsis. Genetics, 175: 1441-1450.

11. Neale DB and Savolainen O (2004). Associated genetics of complex traits in conifers. Trends in Plant Science, 9: 325-3.

12. Rohde A and Bhalerao RP (2007). Plant dormancy in the perennial context. Trends in Plant Science, 12, 217-223.

2. Physiological basis for high yield in bioenergy *Populus*

R Ceulemans, AM Rae, NR Street, MPC Pinel, N Marron, SY Dillen L Ricciotti, F Salani, G Taylor, M Sabatti, G Scarascia Mugnozza, C Bastien

2.1 Introduction

Understanding the complex nature of stemwood yield and the limitations imposed by climate and disease

Hardwood trees provide a sustainable source of carbon-neutral bioenergy able to combat the effects of rising atmospheric CO_2 and other greenhouse gases [1, 2, 3, 4]. *Populus* hybrids are superior hardwood trees for biomass production [5, 6] with high photosynthetic capacity [7], rapid juvenile growth [8] and a high proportion of woody content in the first growing season [9], producing yields of up to 35 Oven Dried Tonnes per hectare per year (ODT ha^{-1} yr^{-1}) [10]. The discrepancy between such potential yields seen in small scale experiments, and yields in large commercial plots is often great, so there is a requirement to better understand the genetic nature and environmental input for yield, enabling breeding programmes to benefit from marker assisted selection and other biotechnological approaches [11, 12].

Populus is now well recognized as the 'model' forest tree [13, 14, 15]. It has several advantages as a model system, including the availability of several linkage maps and pedigrees, transformation systems, and a large genomic resource including EST collections and *Populus* microarrays [16]. Moreover, the release of the *Populus* genome sequence was the first for a tree species (http://genome.jgi-psf.org/Poptr1/Poptr1.home.html). Not only does this facilitate significant advances in our understanding of economically important traits unique to woody plants (such as dormancy and wood formation), it increases our ability to answer fundamental questions in forest ecology and evolution.

Improving the commercial biomass yield of hybrid *Populus* is an essential requirement for an economically viable alternative to other land uses [17] and yet these technologies have to be

fully integrated into most breeding programmes although their potential is significant [12, 18]. Tree height, basal area and number of sylleptic branches are known to be good indicators of final biomass yield in *Populus* [19, 20, 10, 21, 6] as is the ability of the main stem to undergo rapid initial growth with the development of large leaf area in a reasonably short time [22, 23].

Sylleptic branches (late season branches produced from current year buds without a period of dormancy) are important as they possess 50-100% greater leaf area than the leaves on the main stem, allow optimised light interception and inhibit competition by weeds [24, 25]. Production of sylleptic branches in many species has been shown to be highly variable in differing environmental conditions [26]. Influences such as solar radiation, water availability, nutritional level and competition may have a large effect on final biomass production and could potentially be of more importance than genotypic effects. Identification of the genetic components for biomass related traits and their environmental interactions will be important for the production of high yielding genotypes.

Understanding the genetic basis of superior yield and G X E interactions in Populus

Quantitative variation of such complex traits can be dissected by mapping Quantitative Trait Loci (QTL). QTL mapping is a powerful tool to understand the genetic control of complex traits such as biomass, and has been used to look at a number of traits in *Populus* such as rust resistance [27, 28], canopy structure [29], leaf traits [30, 31, 32], stem growth [33, 34], growth rates [35], bud dormancy [36] and metabolites [37]. QTL have also been mapped in the closely related genera, *Salix*, for growth related traits [38], bud dormancy [39, 40], and pest tolerance [41].

However, an understanding of the genetic control of the differential expression of phenotypic traits across environments or genotype x environment (GxE) interaction is of great importance for quantitative genetics and its applications in breeding, conservation, and theory of evolution. Changes in soil water relations [42, 43], cold hardiness [44], soil type [45, 46] and latitude [47] have all been shown to have an effect on biomass traits. The majority of GxE interactions affect traits with low heritability normally at the individual level [48]. Understanding the interaction between the genetic material present in a population and the environment in which that population is grown allows breeders to achieve impact from their

breeding programs [49]. Whether the breeding aims are to select genotypes best suited for a particular environment or to identify genotypes stable across several environments, it is necessary to understand the genetic nature of the GxE interactions.

Genotype x environment (GxE) stability statistics can be used to assess the performance of a pedigree by providing estimates of each genotype's contribution to the GxE complex. One of the most important questions of a stability parameter is whether it is genetic [50]. It can be argued that there are two possible genetic mechanisms underpinning stability [51, 52]; the allelic sensitivity model which implies that the gene is regulated itself in direct response to the environment by the activation of different alleles in different environments; and the gene regulation model where one or more regulatory loci are under the direct influence of the environment and the gene for the trait of interest is switched on or off by the regulatory gene. The collocation of QTL exhibiting GxE interactions and QTL for stability parameters would support the allelic sensitivity model whereas QTL for stability parameters detected in regions other than those for the trait would support the regulatory model [53, 52].

As a starting point we must understand what *morpho-physiological, structural* and *biochemical* traits determine yield in *Populus*. Production physiology suggests that in intensive forest systems, light interception and conversion efficiency are critical determinants of stemwood productivity [54]. Identifying the complex physiological traits associated with yield has been studied extensively with the following associated with high yield in *Populus:*

- Large leaves
- Rapid development and long duration of high LAI
- Increased branchiness and no. of sylleptic branches (those formed from current year meristem in the axils if current year leaves
- High photosynthetic rates and respiration rates
- Highly branched rooting systems.
- Stomata responsiveness.
- Erectophile leaves and branches

Although of value, these data were collected in a multitude of different genotypes and here attention has been focussed on the physiological basis of yield in segregating pedigrees, enabling these traits to be linked to molecular genetic markers, as has been achieved in crop plants such as maize and rice and is described here in Chapter 3. In addition, there are few

studies that have utilised identical plant genotypes in both large-tree field studies and smaller glasshouse trials and controlled environments and very few that have considered the link between physiological traits, QTL across several pedigrees and environments and candidate genes. This is an important omission, since a physiological trait used in marker-assisted selection must be robust and selectable within large progeny numbers and across a wide range of material.

Here we have measured these traits in a wide range of poplar genetic material in mapping populations. Linked to this we have developed a database for input of all raw data, enabling further interogation and analysis. Also, we looked at the phenotypic and genetic variation available in the different full-sib families in terms of aboveground biomass production; we estimated the influence of the genetic background and of the environmental conditions on the relationships between biomass and its predictors. We determined whether one predictor was sufficient to accurately predict tree biomass and whether the use of predictive equations could be standardized among poplar hybrids resulting from various intra- as well as interspecific crosses, and among contrasting environments.

2.2 Materials and Methods

Assessment of stem, canopy and leaf traits for yield

Key traits contributing to yield were measured during year 1 and 2 (confirmed traits) and final coppice measurements were confirmed in year 3 and this was followed by a destructive harvest of all populations and at all sites (with some exceptions for the UK site as detailed below). A summary of the measurements made in given in Table 2.1.

Table 2.1 The detailed physiological traiting - measurements and responsible groups, where POP1, 2, 3, 4 represent *Populus* mapping pedigrees and P01 – P06 are partners within the project. Blocked areas refer to measurements not made.

		Site	Headley (United Kingdom)				Ardon (France)				Cavallermaggiore (Italy)			
Period	Trait	Pedigree	POP1	POP2	POP3	POP4	POP1	POP2	POP3	POP4	POP1	POP2	POP3	POP4
First year	**Stem height**		**P01**		**P04**	**P06**	**P01**	**P03**	**P04**	**P06**	**P01**	**P04**		**P06**
	Stem circumference													
	Number of sylleptic branches													
	Height of the highest sylleptic branch													
	Bud set		**P01**								**P06**			
	Bud burst						**P03**							
Second year	Stem height		P01			P01	P03				P06	P04		P06
	Stem circumference													
	Biomass						P01 + P03 + P04 + P06				P04 + P06			
	Number of resprouts		P01		P04	P06	P01	P03	P04	P06	P01	P04	P04	P06
Third year	Number of sylleptic branches													

In addition to these traits, a number of additional traits were also agreed, that were not measured in all populations at all sites, but were developed as part of intensive experimental campaigns where either a single population or a single site was the centre of focus. These included ecophysiological measurements of ^{13}C carbon discrimination, leaf production rate, cell production and cell expansion rate, photosynthesis and stomatal conductance.

Developing a workplan for physiological traiting

Physiological parameters to be measured in the field were adopted during a meeting in Antwerp in May 2003. At this workshop it was decided that (i) bud set and bud burst would be followed for the first growing season by local teams of the 3 sites (Southampton, UK; Cavallermaggiore, I and Ardon, F); sylleptic branches would be counted at the end of the first growing season; (iii) plant height, stem circumference and height of the highest sylleptic branch would be determined at ends of the first and second growing seasons; (iv) above ground biomass would be weighed at the end of the second growing season; (v) shoots would be counted during the spring of the third year. For all 3 sites, parameters corresponding to points (ii) to (v) would be measured by the Southampton-team (partner 01) for POP1 pedigree, the Antwerp-team (partner 04) for POP3a and POP3b pedigrees, and the Viterbo-team (partner 06) for POP4 pedigree. POP2 pedigree measurements will be shared between partner 01 for the Southampton site, partner 04 for the Cavallermaggiore site and partner 03 for the Ardon site. A full summary is shown below

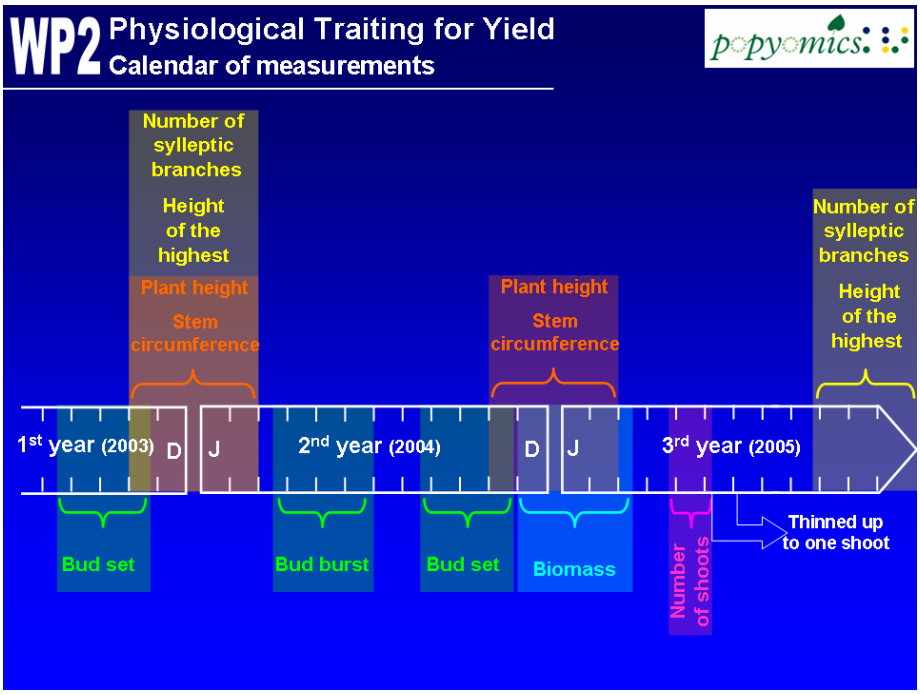


Figure 2.1 A timetable for physiological traiting during the experimental campaigns of year 1, 2 and 3.

During January-March 2004 the four selected traits related to yield were measured at the end of the first growing season (*i.e.* stem height, stem circumference, number of sylleptic branches and height of the highest sylleptic branch) were measured. Traits were estimated for all live plants (maximum 6 replicates) of all pedigrees including the bridge trees in the experiment.

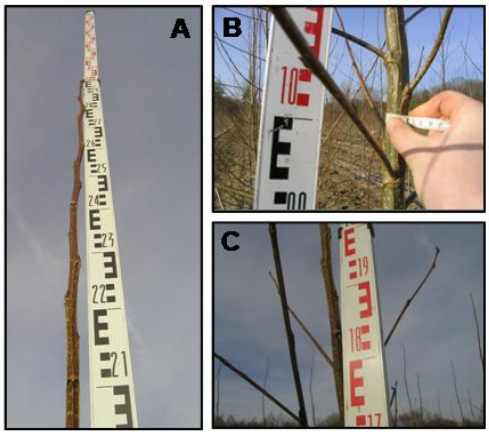


Figure 2.2 F

Assessment of total height (A), circumference at one meter above ground level (B), and height of the highest sylleptic branch (C) for plants of the POP3 pedigrees in Ardon (France) in February 2004

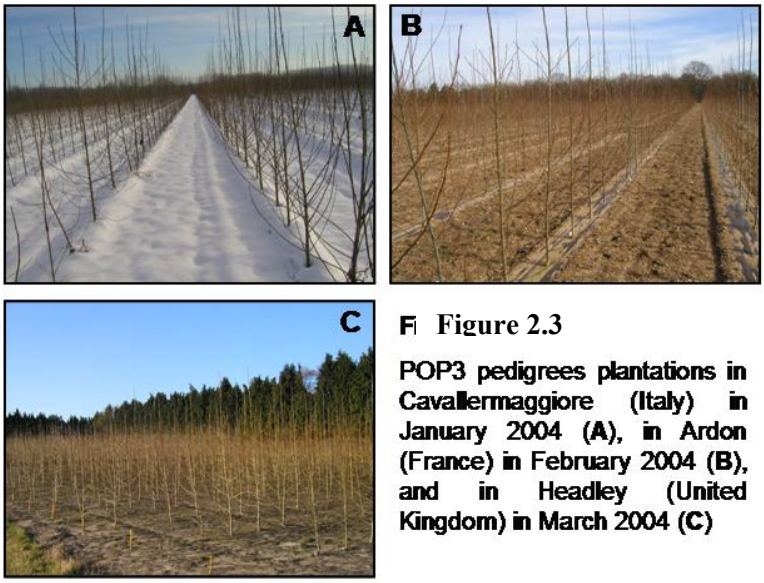


Figure 2.3

POP3 pedigrees plantations in Cavallermaggiore (Italy) in January 2004 (A), in Ardon (France) in February 2004 (B), and in Headley (United Kingdom) in March 2004 (C)

In November 2004, total stem height and stem circumference were measured. Moreover, the general health state of each tree was estimated in order to evaluate its relevance to be harvested. In May 2005, number of resprouts was counted. At the same time, the stumps were cut back, and only the biggest shoot was left on the stump. In January 2005, a last field campaign was undertaken prior to definitive removal of the plantations.

Two campaigns to assess bud phenology

Bud set was measured at the end of the first growing season in all populations, autumn 2003, in the UK and Italian sites. During Spring 2004 bud flush was measured at all sites and for all populations. During autumn 2005, at the end of the third growing season, bud set was monitored for the five pedigrees at the French site, for POP2, POP3a, POP3b and POP4 at the Italian site, and for the LD population at the Belgian site. It was agreed to measure bud set of all trees in Orléans (France) and Geraardsbergen (Belgium). In Cavallermaggiore (Italy), however, all trees of the POP4 pedigree and only a subset of genotypes for the POP2, POP3a and POP3b pedigrees were screened due to the much taller trees in Italy, compared with the other sites, to avoid difficulties in reaching the apical buds. The scale used to score bud set was drawn up by Antje Rohde and Tom Ruttink (P02). Over several weeks (August-October 2005), bud set was scored every two or three days at the three sites and for all selected trees. The scale was divided into six stages from 3 to 0, with score 3 meaning a fully growing shoot, and score 0 meaning a completely closed red-coloured bud (Figure 2.4).

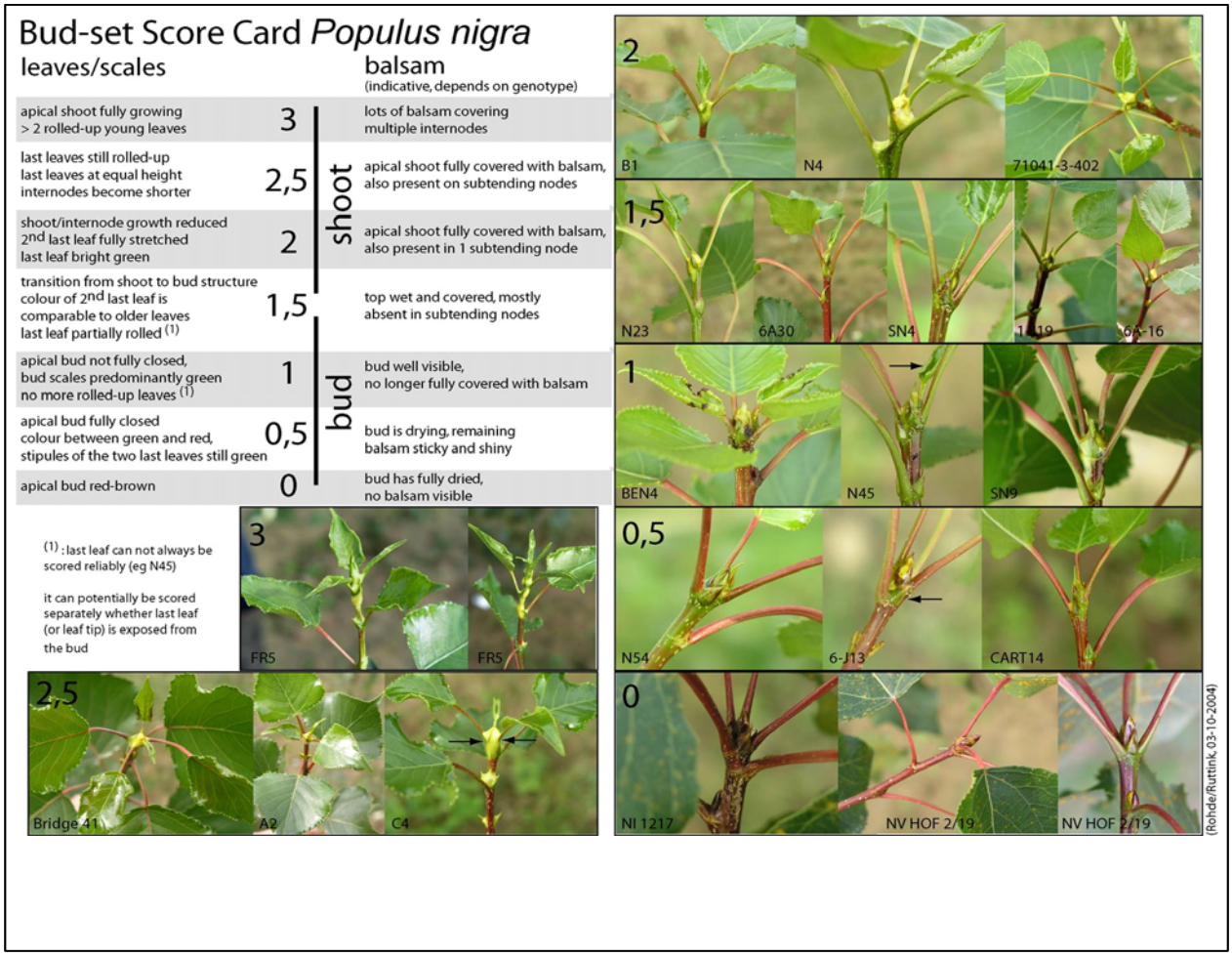


Figure 2.4 Score card describing the seven bud-set stages used to monitor bud closure during fall 2005 in France (Orléans), in Italy (Cavallermaggiore) and in Belgium (Geraardsbergen). The method was developed by Antje Rohde and Tom Ruttink

Statistical handling of the data

Data management and statistical analyses for stem traits were performed with R software (version 2.0.1, A Language and Environment Copyright, 2004). Micro-environmental effects within field sites were minimised using Papadakis spatial correction [55], based on a 7x3 grid on individual data, for POP1 as shown below and using block analysis for the other POPs.

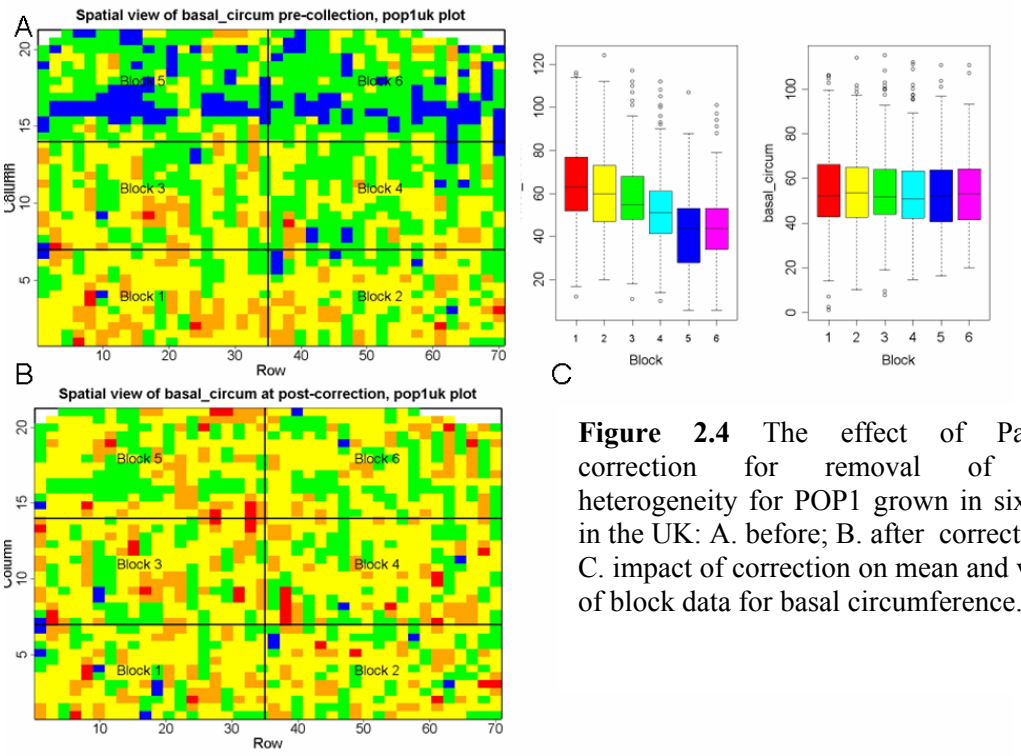


Figure 2.4 The effect of Papadakis correction for removal of spatial heterogeneity for POP1 grown in six blocks in the UK: A. before; B. after correction and; C. impact of correction on mean and variance of block data for basal circumference.

Lines for which there were less than 3 replicates were removed from all further analysis. ANOVA were carried out for all traits scored at each site using R-script and the following model of ANOVA:

$$Y_{ij} = \mu + B_i + G_j + \varepsilon_{ij}$$

where μ is the general mean, B_i is the effect of block i considered as fixed after correction with Papadakis, G_j is the effect of genotype j considered as random, and ε_{ij} is the error. No significant block effects were seen.

Between sites comparison was tested using two-way ANOVA:

$$Y'_{jkl} = \mu + G_j + S_k + GxS_{jk} + \varepsilon_{jkl}$$

where Y'_{jkl} are individual values adjusted for micro-environmental effects using Papadakis' spatial correction, μ is the general mean, G_j is the genotype effect (random), S_k is the site effect (random), GxS_{jk} is the genotype by site interaction (random) and ε_{jkl} is the error.
Variance components were calculated by equating observed mean squares to expected mean squares in a random model [56], and individual broad-sense heritability (H^2) and genotype

heritability (H^2_c) were calculated for each trait (1) at each site using $H^2 = \frac{V_G}{V_G + V_E}$ and $H^2_c = \frac{V_G}{V_G + (V_E/r)}$, and (2) across sites as

$$H^2 = \frac{V_G}{V_G + V_S + V_{GxS} + V_E} \text{ and } H_c^2 = \frac{V_G}{V_G + V_S + V_{GxS} + \left(V_E \big/ r\right)}$$

where V_G is the genetic variance, V_E is the residual variance, V_S is the site variance, V_{GxS} is the genotype by site interaction variance, and r is the average number of replicates. Standard errors of heritability were calculated using the method suggested by [57]. Changes in genotype ranking were tested for using Spearman rank coefficient on genotype means at each site.

Genetic correlations between traits (r_g) were calculated from the variance-covariance matrices obtained from the MANOVA as $r_g = \frac{Cov_{G(x,y)}}{\sqrt{(\sigma^2_{G(x)} \times \sigma^2_{G(y)})}}$, where $Cov_{G(x,y)}$ is genetic covariance between traits x and y, estimated by equating the mean co-products with their expected values according to the Henderson III procedure.

Multi-site analysis

Phenotypic plasticity can be defined as an environmentally-based change in the phenotypic expression of a genotype [59]. In case of significant GxS interaction effects, the stability of genotype ranking across sites was evaluated using the Spearman rank coefficient on genotypic means for each site. Wricke's (1962) ecovalence measures the relative participation of a genotype or a site in the total GxS interaction (W_i and W_k, respectively). The relative contribution of each site in the GxS interaction was calculated by the relative ecovalence of the site $W_k r = \Sigma_j (GxS)_{jk}^2 / \Sigma_{jk} (GxS)_{jk}^2$ for a given genotype j evaluated in k sites. To estimate the genetic variation of the stability parameters through a multi-parametric approach the additive main effect and multiplicative interaction (AMMI) analysis [60, 61] was performed. The goal of the analysis is to summarize the interaction sum of squares (SS) with a few singular value decomposition (SVD) axes, leaving a reduced model with residuals containing mostly noise. Only the early interaction axes have a biological interpretation [62]. The units for the grand mean (μ), the genotype mean deviations (α_g), the environment mean deviations (β_e), the interaction residuals (θ_{ge}), the AMMI residuals (ρ_{ge}) and the singular vector value for

SVD axis n (i.e. number of SVD axes retained in the model) (λ_n) are in the same units as the response Y (stem circumference at the end of the second growing season) [63].

Relationships between traits

Linear correlations between traits were assessed with Pearson's correlation coefficient on a genotypic mean basis. Multivariate analyses using principal component analysis (PCA) were also performed. Only original traits that were not obtained from other traits were included in the analysis. The traits were standardized, and orthogonal factors (i.e., PC1 and PC2 axes) were successively constructed as linear combinations of these traits to maximize variability explained by these factors. Variables were first represented on the plane defined by the two main factors of the PCA; their coordinates were their linear correlation coefficients (Pearson coefficient) with these factors. PCAs were realized for each pedigree and site separately.

Establishment of a database for the physiological data

The **'PHYSIO-TRAIT' database** was structured and tested. The raw data collected at the three sites and for the five pedigrees between autumn 2003 and autumn 2005 were uploaded. Six categories of traits are available: **1.** Tree dimensions, year 1 (stem height and circumference), **2.** Sylleptic branchiness (number and distance to the top of the stem), **3.** Phenology (budset and budburst), **4.** Tree dimensions, year 2, **5.** Biomass production (tree dry weight, realised biomass production, ratio stem biomass / total biomass), and **6.** Resprouting (number of shoots after coppicing). The database is being ordered and will be available to the EVOLTREE network of excellence for tree genomics by December 2007 and then publicly available by 2009.

Other traits and trait description

Other architectural and phenological traits were estimated or calculated as described in the following table.

Table 2.2 Additional traits measured during the project

Traits	Abbreviations	Techniques	Calculation	Unit
Tree dimensions				
Stem height	Height1; Height2	Graduated height pole		cm
Stem circumference at 1 m above ground level	Circum1; Circum2	Taper		mm
Ratio stem height / stem circumference	Ht_Cc1; Ht_Cc2		Stem height / stem circumference	cm mm^{-1}
Stem height increment	DeltaHeight		Stem height year2 - stem height year1	cm
Stem circumference increment	DeltaCircum		Stem circumference year2 - stem circumference year1	mm
Architectural traits				
Number of sylleptic branches	Syllep1	Counting		
Distance to top of sylleptic branches	Syltotop	Graduated height pole		cm
Number of resprouts	Shoots	Counting		
Phenology				
Bud flush	Budburst	bud closed/opened (Van Slycken 1995)		Day of year
Bud set	Budset1	bud opened/closed		Day of year

Destructive harvest and developing equations to describe biomass production

Estimation of tree dimensions and number of sylleptic branches

The total number of sylleptic branches (i.e. branches developed from axillary buds not undergoing a dormant period (64; Syllep1) per stem was counted at the end of the first growing season (i.e. winter 2003–2004) for all trees at the three sites. At the end of the second growing season (i.e. winter 2004–2005) stem height (Height2) of each tree was measured to the nearest cm using a graduated height pole and circumference (Circum2) was estimated to the nearest mm at 1 m above the ground using a tape.

Harvest procedure

In December 2004 and January 2005, at the end of the second growing season, trees were harvested at the three sites for the POP1 and the POP4 families, and at the French and Italian sites only for the POP2, POP3a and POP3b families. Due to the high susceptibility of the F_1 hybrids between *P. deltoides*, *P. nigra* and *P. trichocarpa* to foliar rust, the POP2, POP3a and POP3b families were not sampled at the UK site. A sample of 20 genotypes representative of the genetic variation in growth and syllepsis was selected by optimization with the following quality criterion (QC):

$$QC=[Var(Circum2_{sample})-Var(Circum2_{population})]^2+[Var(Syllep1_{sample})-Var(Syllep1_{population})]^2+[Cov(Circum2-Syllep1_{sample})-Cov(Circum2-Syllep1_{population})]^2$$

where Circum2 = stem circumference at 1 m above the ground level at the end of the second growing season and Syllep1 = number of sylleptic branches counted at the end of the first growing season. The QC was compared for 10,000–20,000 random samplings of 20 genotypes by using a procedure developed by INRA Orléans (France) using R script (version 2.0.1., A Language and Environment Copyright, 2004).

Three replicates of each of the 20 selected genotypes showing the closest performance to the genotypic mean (based on the six planted replicates) in terms of stem circumference and number of sylleptic branches were harvested. Trees were cut at 5 cm above the ground and fresh weight of stems and branches was measured separately. Samples were randomly taken as follows: one from the first-year stem growth, one from the second-year stem growth and two branch samples corresponding to at least 10% of the total fresh branch weight. All samples were weighed before and after being dried at 105 °C in a drying oven during 5 days for determination of dry/fresh weight ratios. In France, however, all branches were used to determine the branch dry/fresh weight ratios.

Data analyses

Data management and statistical analyses were performed with R software. Site and genotypic means were calculated with their standard error. In case of non-normality of the distribution of residuals, the Box–Cox procedure was performed and the lambda value of this method was used to transform the data. Independently for each family and site, the following ANOVA model was adjusted to individual data: $Y = \mu + B_i + G_j + \varepsilon_{ij}$, where the effect of the block i (B_i) was considered as fixed and the effect of the genotype j (G_j) as random. Variance components of genetic (σ_G^2) and residual (σ_ε^2) effects were calculated through the least square method [56]. Broad-sense heritability on an individual basis (H^2) was calculated as the ratio between genetic and phenotypic variance (σ_G^2/σ_P^2 where $\sigma_P^2 = \sigma_G^2 + \sigma_\varepsilon^2$) [65] and [57]. The standard error (S.E.) of broad-sense heritability was calculated.

The linear model was used to develop the biomass equations for each family and each site independently as well as for a common equation including all families and sites. Multiple regression analysis was performed in order to look for a subset among a set of potential predictors (Circum2, Height2, and Syllep1) that describes the response sufficiently well. As

the residuals have to be normally distributed, the Box–Cox procedure was performed to search for the most appropriate transformation of the dependent variables of the linear regression model. Firstly, a regression model was fit, and then an optimal transformation was sought for within the Box–Cox power transformations. The model formula was:

$$DW=\beta_0+\beta_1 Circum2+\beta_2 Height2+\beta_3 Syllep1$$

where DW = dry weight of stem (StemDW), branches (BranchDW), or tree dry weight (TreeDW), and the parameters β_0, β_1, β_2 and β_3 were estimated using the method of least squares. One by one, a predictor was removed when the *t* value of the regression analysis was non-significant ($P > 0.05$) (H_0: $\beta_1 = \beta_2 = \beta_3 = 0$). The predictor with the highest non-significant *P* value was removed after which the test was repeated. Thus, only the predictors contributing significantly to the equation were included in the final equations. A simple regression analysis was performed with the predictor Circum2 only. For each equation, the adjusted R^2_{Multi} and $R^2_{Circum2}$ values (determination coefficients of the multiple and simple regression analyses, respectively) were preferred as they are adjusted for number of observations.

Average bias was defined as:

$$\text{Average bias} = \frac{\sum_{i=1}^{n} y_i - \hat{y}_i}{n}$$

Standard error of estimate (S.E.E.) was defined as:

$$\text{S.E.E.} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n - m - 1}}$$

where y_i = measured value; $\hat{y}_i$ = estimated value; n = number of observations; m = number of independent variables in the model. As average bias is defined as the measured value minus the estimated value, a positive number stands for an underestimation while a negative number means an overestimation. S.E.E. is a measure of the spread of the data within the entire

dataset around the regression line or surface. Thus, the lower the value of S.E.E., the better the fit to the data [66].

To test if the regression lines were the same for all families and sites, an ANCOVA model was performed with Circum2, Height2 and Syllep1 as continuous random Due to the unbalanced model (i.e. three sites for the T × D and the A × A families, and only two sites (France and Italy) for the D × T^1, D × N and D × T^2 families), the ANCOVA was first carried out with the five families at the French and Italian sites only. Then, in order to test if there were differences between the biomass equations at the UK site and the two other sites, the ANCOVA was repeated for the T × D and the A × A families at the three sites. After fitting the model to these data, the predicted response from the model for values of these two families at the three sites was verified with the generic 'predict' function in R. The *F*-test of the ANCOVA was performed on the five families at the French and Italian sites only because of the rust infection which affected the growth performances and the allometric relations of most families at the UK site.

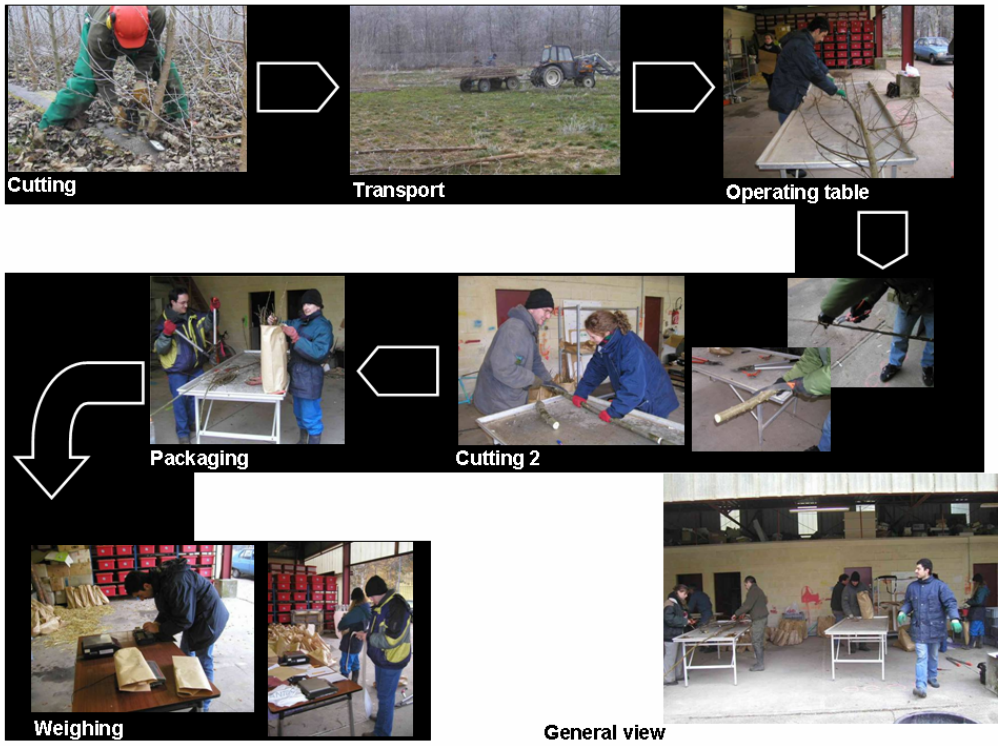


Figure 2.6 A summary of the destructive harvest process from cutting through to weighing where teams from UK, F, B and Italy contributed to the data collection.

2.3 Results and Discussion

Biomass production and tree dimensions

Among the five pedigrees, the F_1 pedigrees involving *P. deltoides* and *P. trichocarpa* (POP2 and POP3b) were the most productive ones (Figure 2.7A). For the POP2 pedigree, values of 10.7 ODT ha^{-1} $year^{-1}$ of realised above-ground wood biomass production ($AWP_{realised}$) were obtained in Italy, and 6.1 ODT ha^{-1} $year^{-1}$ in France. For POP3b, $AWP_{realised}$ was 8.9 and 6.2 ODT ha^{-1} $year^{-1}$, at the Italian and French sites, respectively. On the contrary, the *P. deltoides* x *P. nigra* pedigree (POP3a) and especially the F_2 *P. trichocarpa* x *P. deltoides* one (POP1) were the least performing pedigrees, with $AWP_{realised}$ ranging from 1.8 to 3.0 ODT ha^{-1} $year^{-1}$ and from 3.5 to 5.4 ODT ha^{-1} $year^{-1}$, respectively. For the intraspecific *P. alba* pedigree (POP4), intermediate values of $AWP_{realised}$, ranging from 3.2 to 7.6 ODT ha^{-1} $year^{-1}$, were observed. It is noteworthy to precise that there was no biomass estimation in UK for the POP2, POP3a and POP3b pedigrees due to a severe rust attack (*Melampsora larici-populina*) which drastically affected the productivity of these three pedigrees at this site during the second growing season. The hybrids from the cross between *P. deltoides* and *P. trichocarpa* are indeed particularly sensitive to rust. However, in the absence of pathogens, hybrids from this cross are known for their very vigorous growth, probably owing to the relative proximity to their respective ancestor botanical sections (*Aigeiros* and *Tacamahaca*, respectively) [51]. In contrast, *P. nigra* is a pioneer species, conferring on its progeny high tolerance to environmental constraints at the expense of growth performance.

During the project period numerous growth and biomass traits were determined (see Table 2.2) in Materials and Methods). Among these traits, stem circumference estimated at the end of the second growing season (Circum2) was the most closely related trait to biomass production for the five pedigrees at the three sites. Indeed, according to the multiple regression analysis, Circum2 was the best predictor of biomass production ($0.75 < R^2 < 0.97$; data not shown) as compared to the stem height estimated at the end of the second growing season (Height2) and to the number of sylleptic branches (Syllep1). Circum2 is an interesting trait to be considered as a surrogate of growth also due to more important genetic variation, i.e. moderate to high values of individual values of broad-sense heritability (H^2_{ind}, 0.10 - 0.66) and coefficient of genetic variation (CV_G, 7% - 35%) (Figure 2.7B). As for the biomass

production, the ranking of the five pedigrees for Circum 2 was POP2 > POP3b > POP4 = POP3a > POP1 at the three sites. Heritability values and percentage of phenotypic variation assigned to the genotype were higher for the traits scored at the end of the second growing season than at the end of the first one, possibly due to a C-effect during the first growing season, i.e. physiological preconditioning due to differences among stock plants of the same genotype. Tree performances recorded during the first growing season have to be handled carefully because they are not necessarily good indicators of the future growth of the trees.

More precisely, concerning the extent of the genetic variation within pedigrees, H^2_{ind} values for Circum 2 were moderate to high (0.10 – 0.60), except for the F_2 hybrids of the POP1 pedigree ($H^2 = 0.65$). Among the F_1 hybrids, the POP3a pedigree showed the highest values of H^2_{ind} (around 0.50). The genetic variation was the highest for the POP1 pedigree (CV_G = 30%) and the lowest for the POP2 one (CV_G = 8%). Positive and high heterosis (or hybrid vigor) was found for the F_1 hybrids, especially for the hybrids of the POP2 and POP3b pedigrees. Heterosis of the F_1 hybrids of the *P. deltoides* x *P. trichocarpa* crosses is due, in part, to the combination of rapid height growth of *P. trichocarpa* with the rapid diameter growth of *P. deltoides* [67]. On the contrary, negative values of heterosis were found for the POP1 pedigree. However, estimation of hybrid vigor has to be treated carefully as the effect of competition may be important on the performance of the parents growing with their progeny.

In terms of site influence, the *P. trichocarpa* parents showed a higher productivity at the UK site while the *P. nigra* parents performed better in warmer and drier conditions of Italy (Figure 2.7C). The *P. trichocarpa* species is indeed adapted to a typical oceanic climate (mild and wet); in contrast, the *P. nigra* species is a Mediterranean species (the origin of which is in Iran) [9, 68, 69]. The growth of the *P. alba* parents was dependent on the genotype, the two parents of the POP4 pedigree showing very variable values. During the second growing season, all hybrids of the five pedigrees were more productive under the more favorable conditions of Italy (i.e. loamy soil, and warmer and drier conditions as compared with France and UK). This is in line with the growth of all pedigrees during the first growing season, except for the POP3b pedigree which showed a better growth in UK as compared to Italy during the first growing season only. So, for this pedigree, the preferences of the parents in terms of climate may influence the response of hybrids to the different sites during the

establishment year only. Afterwards, the quality of the local conditions could become the dominant factor in determining the growth of the hybrids.

Tree architecture

Crown architecture and canopy density of trees are intimately related to stand productivity. Crown architecture determines leaf display, leaf distribution and canopy density, and therefore influences light interception [70]. It has been demonstrated that the main factor giving rise to the high leaf area of poplar are the sylleptic branches (i.e. branches resulting from the development of newly initiated lateral axes without the apical meristem of those axes passing through a dormant period). Moreover, sylleptic branches translocate a larger proportion of carbon to stem than proleptic branches, and their contribution to stem growth is consequently primordial [24]. The resprouting vigor of the plants after coppicing is also a factor well-known to be strongly linked to the whole productivity of the plant.

Sylleptic branchiness was affected by both site conditions and pedigrees. The highest numbers of sylleptic branches were produced by the F_1 pedigrees involving *P. deltoides* and *P. trichocarpa* (POP2 and POP3b) and by the *P. alba* pedigree (POP4), while the F_2 *P. deltoides* x *P. trichocarpa* (POP1) and F_1 *P. deltoides* x *P. nigra* (POP3a) pedigrees showed the lowest number of sylleptic branches. Differences between pedigrees also concerned the pattern of the establishment of the sylleptic ramification. Indeed, for the POP2, POP3b and POP4 pedigrees, the sylleptic production started earlier during the growing season (lower along the stem) and stopped later (higher along the stem) than for the POP1 and POP3a pedigrees. Due to the very high number of sylleptic branches produced by most of the male parents (notably the three *P. trichocarpa* genotypes and the *P. alba* genotype), heterosis (hybrid vigor) was generally low. For all pedigrees, a quite large range of phenotypic variation was observed, ranging from 0 to 56 branches per tree, on average. Individual values of broad-sense heritabilities were moderate to high for the POP1 pedigree (ranging from 0.52 to 0.59) and low to moderate for the other four pedigrees (ranging from 0.16 for POP3b in UK, to 0.53 for POP3a in Italy). In the same way, the POP1 pedigree exhibited high coefficients of genetic variation (from 33.3% to 44.9%) as compared with the other pedigrees, for which the values of CV_G ranged from 14.4% (for POP3b in UK) to 33.2% (for POP3a in France). In most cases (except for the POP3b pedigree), Italian conditions were the most

favorable for the development of the sylleptic ramification. The highest values of broad-sense heritability were observed in Italy (except for the POP1 pedigree) while the highest values of coefficients of genetic variation were observed in France or in UK for the five pedigrees.

The highest numbers of resprouts after coppicing were observed for the POP3a and POP4 pedigrees. As for the number of sylleptic branches, broad-sense heritability and coefficients of genetic variation values were high for the POP1 pedigree. However, low values were observed for the other four pedigrees (H^2_{ind} ranging from 0.06 to 0.40 and CV_G ranging from 7.4% to 31.2%). These low values as compared with the ones observed for the number of sylleptic branches, associated with the fact that one of the less productive pedigrees (POP3a) showed the highest numbers of resprouts after coppicing, limit the relevance of the use of this trait as indirect indicator of the future productivity of the trees.

Phenology

The seasonal cycle of growth and dormancy is a distinct character of perennial plants and represents one of the most basic adaptations of trees to their environment. The recurrent transitions of meristems into and out of dormancy are of primary significance to plant productivity and survival. These transitions are tightly linked to the yearly dates of bud burst and bud set that in turn delimit the growing season [72].

Bud set was screened at the end of the first growing season after establishment of the plantations in UK and Italy (2003), and at the end of the first growing season after coppicing in France and in Italy (2005; data not shown). Bud burst was measured at the three sites at the beginning of the second growing season (2004). Between pedigrees, only a few variation was recorded for average date of bud set measured in 2003 (3 days), except for the POP4 pedigree. Indeed, the *P. alba* pedigree displayed the latest bud set among the five pedigrees (between 10 and 15 days later than the other pedigrees, on average), possibly due to the inheritance of the very late bud set of the male parent *P. alba* '14P11' as compared with the other parents. Among the other parents, the bud set was slightly earlier for *P. nigra* 'Ghoy' than for the other parental genotypes, explaining the slightly more precocious bud set of the *P. deltoides* x *P. nigra* pedigree (POP3a) as compared with the crosses involving *P. trichocarpa* (POP1, POP2 and POP3b). Very low values of coefficients of genetic variation were observed for all

pedigrees and both sites, ranging from 0.7% to 3.1%. Values of heritability were moderate (ranging from 0.13 to 0.56), and significantly lower values were found in UK for the POP2 and POP3b pedigrees possibly due to their susceptibility to rust. It is important to precise that the absolute comparison of the data between the UK and Italian sites was unfortunately not possible because of different criteria used to estimate the dates of bud set at the two sites.

In *Populus*, the date of bud burst is one of the most readily mapped quantitative traits, and 98% of the phenotypic variance is under genetic control. As for bud set, bud burst is also an adaptive trait that shows latitudinal and altitudinal trends within species. Among the five pedigrees, the POP3b pedigree showed the earliest bud flush in 2004 (day 89) as the POP1 pedigree showed the later one (day 94). Individual values of broad-sense heritability were high at the Italian and French sites (ranging from 0.55 to 0.86) for all pedigrees and low to moderate at the UK site (ranging from 0.25 to 0.52). As for the bud set, the coefficients of genetic variation calculated within sites were very low for all pedigrees and sites, ranging from 1.5% to 7.1%. However, the multi-site analysis revealed a high genetic effect for the POP2, POP3a and POP4 pedigrees (from 27% to 38%). Among the parental genotypes, *P. trichocarpa* showed the earliest bud burst as compared with *P. deltoides*, *P. nigra* and *P. alba*. With the exception of the F_2 pedigree (POP1), this property conferred a slightly earlier bud burst to the interamerican pedigrees (POP2 and POP3b) as compared with the POP3a and POP4 pedigrees.

Multi-site analyses

All genotype by site (GxS) interactions concerning stem circumference at the end of the second growing season (Circum2) were highly significant ($P \leq 0.001$) (Figure 2.7D). However, the interaction represented less than 10% for all sites and all pedigrees. In order to analyze the GxS interaction, genotypic stability among sites was estimated through both an **uniparametric approach**, by calculating the Wricke ecovalences [73], and a **multiparametric approach**, by using the Additive Main effects and Multiplicative Interaction (AMMI) model, also called Principal Components analysis of the GxS interaction. Data concerning the AMMI model are presented for the POP2, POP3a and POP3b pedigrees only (Figure 2.8).

The relative site ecovalences, giving a measurement of the relative participation of each site in the GxS interaction, indicated a special response to the Italian conditions for Cirucm2 of the POP1, POP3a and POP4 pedigrees (Figure 2.7D). This was not true for the POP2 and POP3b pedigrees for which the interaction was mainly associated to the UK site, possibly due to the rust attack in UK which has notably affected their productivity. Nevertheless, all pedigrees showed their best performances in Italian conditions. The southern location of the site is indeed characterized by a more favorable climate: high-radiation conditions, higher temperatures and sufficient water supply. Furthermore, soil texture is pure loam while the French and the UK soils are sandier. The relative genotype ecovalences highlighted that only a few number of genotypes were mostly responsible for the GxS interaction (e.g. 66 genotypes for the POP2 pedigree).

Thanks to the AMMI model, two independent components of the GxS interactions of Circum 2 could be identified. These two components (MUL1 and MUL2) explained, respectively, 59.6% and 40.4% of the interaction for the POP2 pedigree, 81.8% and 18.2% for the POP3a pedigree, and 55.1% and 44.9% for the POP3b pedigree (Figure 2.8). The MUL1 components showed a contrast between the UK and Italian sites for the POP2 pedigree, or between the UK and both other sites for the POP3a and POP3b pedigrees, probably explained by the higher growth showed by all pedigrees in Italy as compared with both other sites. For the POP3a pedigree only, a negative correlation between Circum2 and MUL1 was observed, the most productive genotypes being the least stable among sites. The MUL2 components showed a contrast between France and UK for the POP3a and POP3b pedigrees or between France and both other sites for the POP2 pedigree. However, the share of the GxS interaction explained by this component was significant for the POP2 and POP3b pedigrees only (18.2% only for the POP3a pedigree), highlighting the effect of foliar rust in UK to which these two Interamerican pedigrees are very sensitive. For the POP2 pedigree only, a negative correlation between the main effect (Circum2) and MUL2 was observed, the most productive genotypes being the least stable among sites. For the three pedigrees, a wide range of genotypic variation to specific adaptation to the three sites was observed. The relevance of the use of the two multiplicative components for further QTL detection, as indicator of the relative stability of the different genotypes to varying environmental conditions, was also conferred by the normal distribution showed by these stability traits.

In spite of the high significance of the GxS interactions, the Spearman rank coefficients displayed a moderate to good stability of the genotypic ranking for Circum2 at the three sites for the five pedigrees, highlighting the fact that the GxS interactions were mainly due to differences between genotype in magnitude rather than to a general trade-off in performance among sites (Figure 2.7D). Overall, a genotype selected for its growth in one site will remain productive if grown at another site.

Relationships between traits

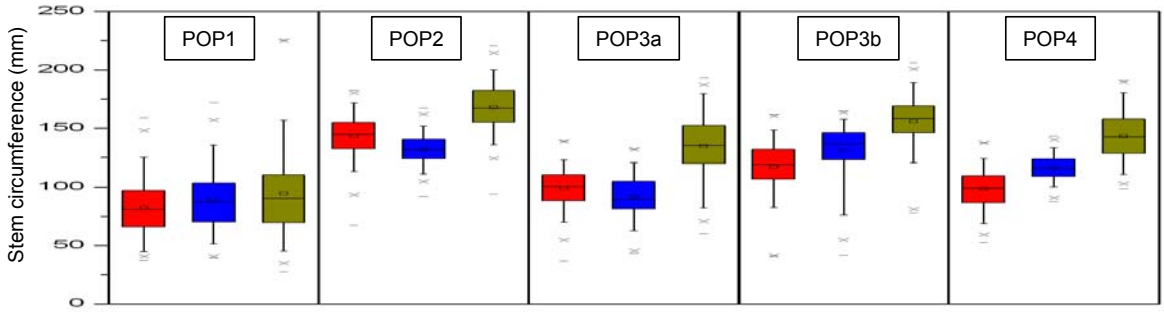
In order to look at the overall relationships between traits, principal components analyses (PCA) were performed for the five pedigrees and for the three sites, independently. Three categories of traits were used: (1) growth traits (Circum1, Circum2 and Biomass; see Materials and Methods for trait abbreviations), (2) phenological traits (Budburst and Budset) and (3) architectural traits (Syllep1 and Shoots). The main planes of the PCA (PC1 x PC2) explained between 58% and 76% of the phenotypic variability, with PC1 explaining between 40% and 59% of the variability (Figure 2.9). For each pedigree and site, the PC1 axes were defined by the traits Circum1, Circum2 and Biomass and could be considered as growth axes. The PC2 axes were defined by the phenological traits (Budburst and Budset) for the POP1, POP2, POP3a and POP3b pedigrees, while the architectural traits (Syllep1 and Shoots) were mostly positively linked to the PC1 axes. For the POP4 pedigree, on the contrary, the PC2 axes were mainly defined by the architectural traits, while negative correlations were observed between Budburst and growth traits along the PC1 axes.

The productivity of the different pedigrees showed various links with phenological and architectural traits depending both on the kind of cross and on environmental conditions. For the four F_1 as well as F_2 interspecific pedigrees generated from the crosses between *P. deltoides*, *P. trichocarpa* and *P. nigra* (POP1, POP2, POP3a and POP3b), tree productivity was tightly linked to the number of sylleptic branches and to the number of resprouts after coppicing (with the exception of the Italian site for the POP1 and POP2 pedigrees) and, on the contrary, independent on the dates of bud set and bud burst (with the notable exceptions of the dates of bud set at the English site only, for the POP2, POP3a and POP3b pedigrees). Therefore, for these four pedigrees, it appeared that, under the more favorable Italian edaphic as well as climatic conditions (loamy soil, warmer and drier climate) as compared with both

English and French conditions, higher tree productivity was attributable neither to higher number of branches or shoots, nor to longer growing season. On the other hand, under the relatively less favorable English conditions (sandy soil, colder and wetter climate), a higher tree productivity was attributable both to a larger number of sylleptic branches and resprouts after coppicing, and to a longer growing season due to a delayed date of bud set. On the other hand, for the intraspecific *P. alba* pedigree (POP4), the productivity was independent on the number of sylleptic branches and on the number of shoots after coppicing as well as on the date of bud set, but negatively and significantly correlated to the date of bud burst. For this pedigree only and irrespective of site conditions, tree productivity was under the influence of the date of onset of the growing season.

Concerning sylleptic branchiness and resprouting, more detailed studies have been realized and published for the POP1, POP3a and POP3b pedigrees. For the POP1 pedigree, sylleptic branchiness has been shown to significantly and positively contribute to growth and biomass production [6, 74]. However, for the POP3a and POP3b pedigrees during the first growing season, no improvement in genetic gain was observed when selection was based on syllepsis, despite moderate values of heritability associated with strong genetic correlations with growth traits [75]. Sylleptic branchiness appeared not to be a relevant criterion for indirect selection for biomass production in these two pedigrees, except in case where growth conditions are favorable for growth and development of sylleptic branchiness. During the second growing season, biomass production within the POP3a and POP3b pedigrees has been observed to be tightly linked to the biomass allocated to branches as well as to the number of resprouts produced after coppicing (Marron *et al.* in preparation). However, the most productive pedigree (POP3b) was also the one producing the least number of shoots after coppicing and allocating the least biomass to branches, highlighting the fact that relationships between biomass production and its potential predictor traits, observed within one pedigree, are not necessarily valid to compare genotypes belonging to different pedigrees.

A. Boxplots



Boxplots of the genotypic means of stem circumference 1m above ground level (mm) at the end of the second growing season for the five pedigrees (POP1, POP2, POP3a, POP3b and POP4) at the three sites (UK (red), France (blue) and Italy (green)). Each box is divided at the median and the height of the box represents the interquartile range. The vertical lines represent the range of values, with outliers indicated by crosses and dashes.

B. Pedigree genetic characteristics

	POP1			POP2			POP3a			POP3b			POP4		
	UK	F	I	UK	F	I	UK	F	I	UK	F	I	UK	F	I
Means (mm)	81.5	89.5	94.0	127.9	119.8	151.2	99.2	92.4	136.6	117.4	131.4	156.0	99.2	116.4	143.7
H^2_{ind}	0.61	0.65	0.66	0.15	0.15	0.12	0.44	0.57	0.39	0.27	0.57	0.27	0.14	0.10	0.26
CV_G (%)	28.0	26.9	35.1	7.9	6.7	9.0	14.5	18.2	17.5	14.7	17.2	12.0	12.4	6.2	11.3
CV_R (%)	23.0	19.7	25.2	18.6	16.3	24.3	14.8	15.9	21.7	16.7	15.1	19.8	31.2	18.2	18.7

Means, broad-sense heritabilities on individual basis (H^2_{ind}), coefficients of genetic variation (CV_G) and residual variation (CV_R) of stem circumference at the end of the second growing season for the five pedigrees (POP1, POP2, POP3a, POP3b and POP4) at the three sites (UK, France (F) and Italy (I)).

C. Parental values

	POP1			POP2			POP3a			POP3b			POP4		
	UK	F	I	UK	F	I	UK	F	I	UK	F	I	UK	F	I
?	146.7	155.3	126.2	61.7	114.3	104.5	68.2	78.5	NA	68.2	78.5	NA	61.7	114.3	104.5
?	111.3	122.0	167.8	87.0	56.3	41.3	87.0	67.8	104.5	122.8	94.4	61.3	87.0	56.3	41.3

Means of stem circumference at the end of the second growing season (mm) for the parents of the five pedigrees (POP1, POP2, POP3a, POP3b and POP4) at the three sites (UK, France (F) and Italy (I)). NA = non available. Mean values calculated from 3 to 6 replicates.

D. Multi-site analyses

	POP1			POP2			POP3a			POP3b			POP4		
1. σ^2_G (%)	48.2			7.7			18.9			15.8			5.8		
σ^2_{GxS} (%)	9.6			4.9			7.1			6.0			4.7		
	UK	F	I	UK	F	I	UK	F	I	UK	F	I	UK	F	I
2. w_i^r (%)	19.2	17.8	63.0	44.6	24.2	31.2	24.8	16.3	58.9	31.3	29.2	39.5	32.1	25.8	42.1
3. $R_{spearman}$	UK	F	I	UK	F	I	UK	F	I	UK	F	I	UK	F	I
UK		0.74 ***	0.64 ***		0.26 ***	0.31 ***		0.71 ***	0.61 ***		0.56 ***	0.37 ***		0.33 ***	0.36 ***
F			0.58 ***			0.35 ***			0.70 ***			0.51 ***			0.32 ***

1. Relative importance of genetic (σ^2_G) and genotype by site (σ^2_{GxS}) effects in the phenotypic variation across the three sites (UK, France (F) and Italy (I)) for the five pedigrees. 2. The relative ecovalences per site for stem circumference at the end of the second growing season for the five pedigrees. 3. Spearman rank coefficients ($R_{spearman}$) based on genotypic means between UK, France (F) and Italy (I) for stem circumference for the five pedigrees. The levels of significance for the Spearman correlations are indicated as: $P = 0.001$ = ***.

Figure 2.7 . Genetic diversity of stem circumference measured at the end of the second growing season for the five hybrid poplar pedigrees and their parents within (A, B and C) and among sites (D).

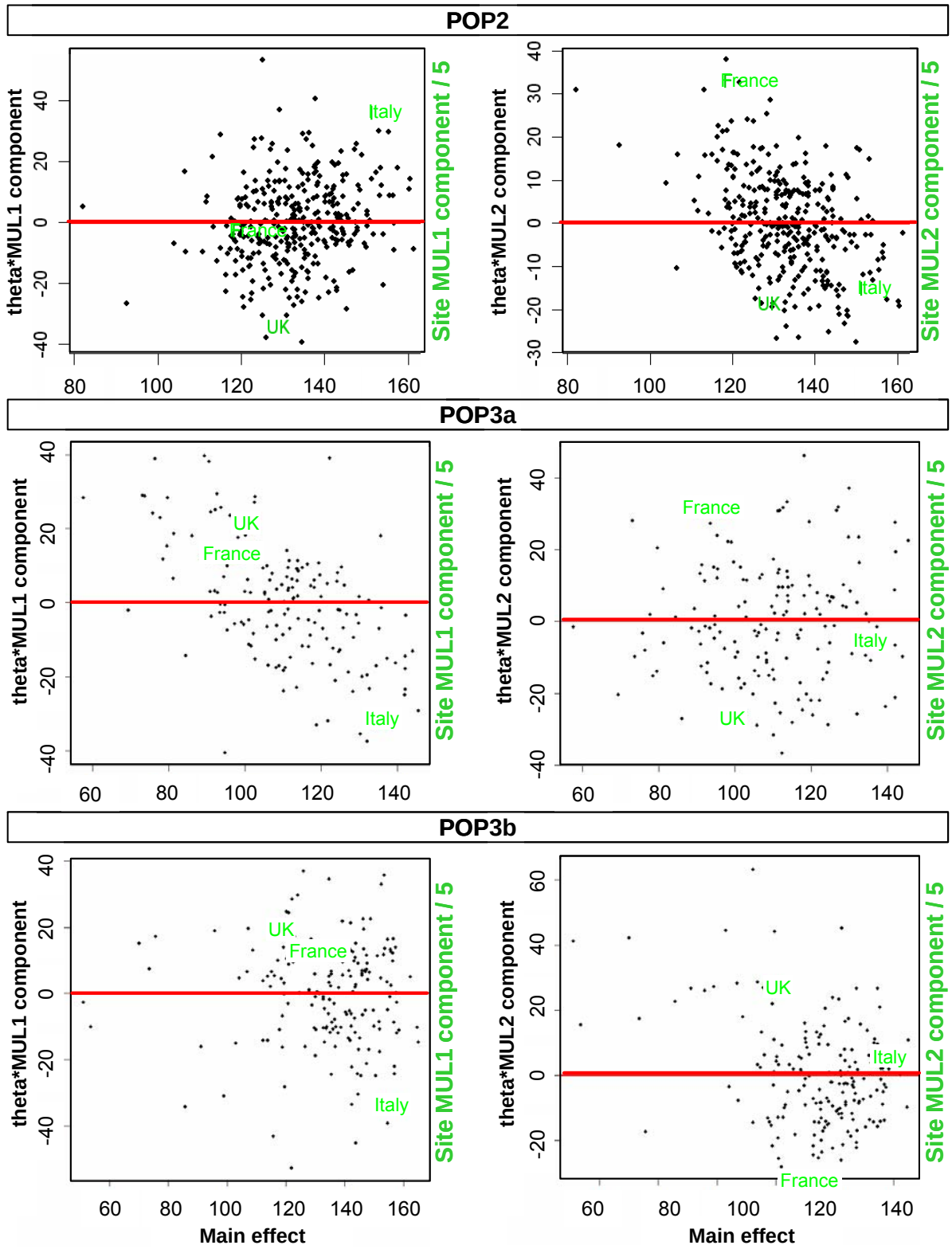


Figure 2.8 Plot of the multiplicative terms (MUL1 and MUL2) given by the AMMI (Additive Main effect and Multiplicative Interaction) model as a function of the main effect (i.e. stem circumference at the end of the second growing season, Circum2, mm) for the genotypes of the POP2, POP3a and POP3b pedigrees, in France, in Italy and in UK. MUL scores close to zero indicate little or no interaction. Displacement along the vertical axis indicates interaction differences between genotypes. Displacement along the horizontal axis indicates differences in genotype main effects.

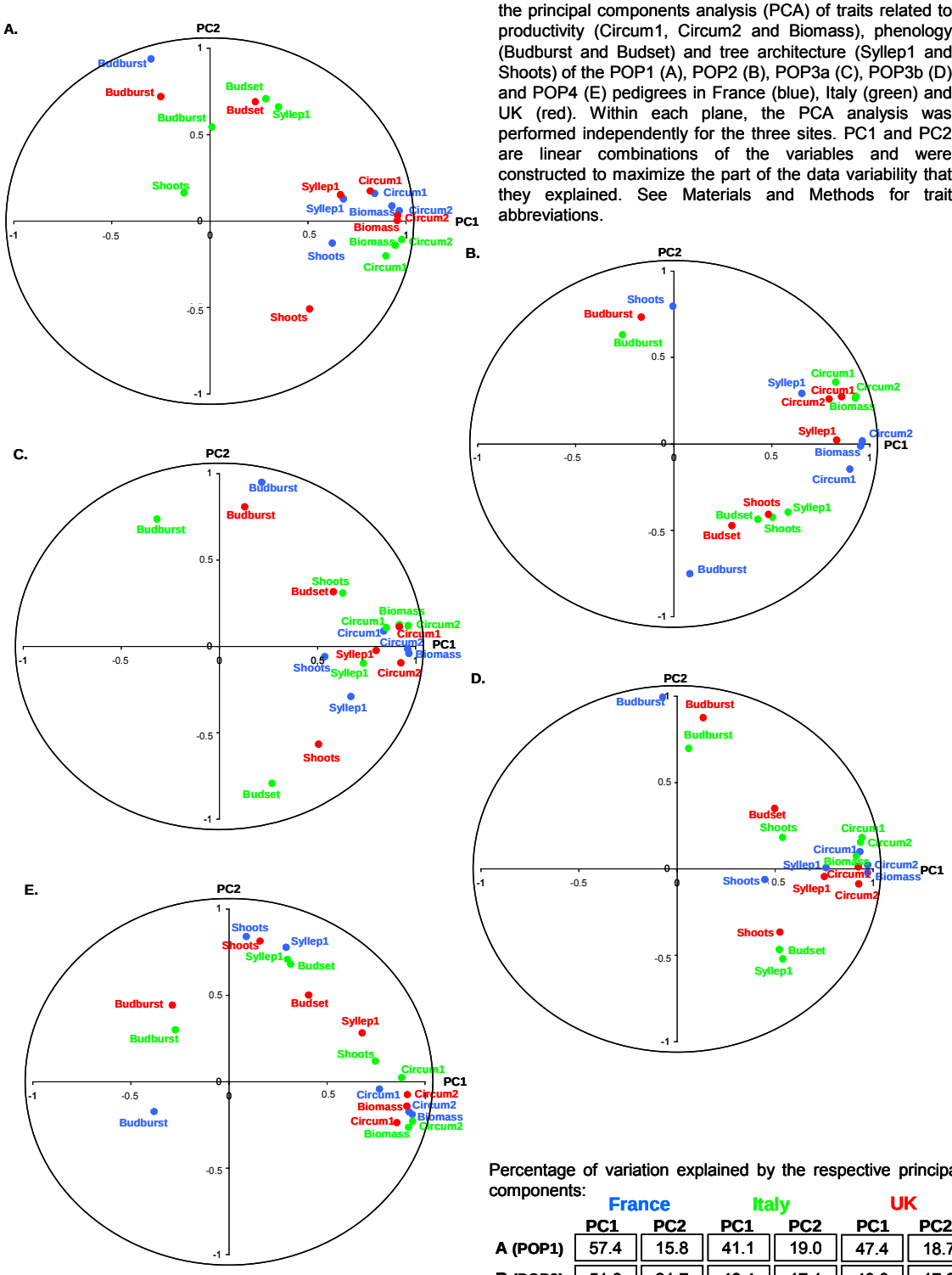


Figure 2.9 Distribution in the factorial planes PC1 x PC2 of the principal components analysis (PCA) of traits related to productivity (Circum1, Circum2 and Biomass), phenology (Budburst and Budset) and tree architecture (Syllep1 and Shoots) of the POP1 (A), POP2 (B), POP3a (C), POP3b (D) and POP4 (E) pedigrees in France (blue), Italy (green) and UK (red). Within each plane, the PCA analysis was performed independently for the three sites. PC1 and PC2 are linear combinations of the variables and were constructed to maximize the part of the data variability that they explained. See Materials and Methods for trait abbreviations.

Percentage of variation explained by the respective principal components:

	France		Italy		UK	
	PC1	PC2	PC1	PC2	PC1	PC2
A (POP1)	57.4	15.8	41.1	19.0	47.4	18.7
B (POP2)	51.3	21.7	46.4	17.1	40.0	17.9
C (POP3a)	55.6	16.7	51.4	18.8	49.0	18.3
D (POP3b)	59.1	16.7	49.4	15.3	47.0	17.2
E (POP4)	42.4	23.5	47.3	17.0	44.6	18.2

2.5 Conclusions

Biomass production is a complex trait, which could be partitioned into functional and structural components. The analysis of productivity components and their genetic control in *Populus* is a crucial step in the early selection process of highly productive genotypes: the breeder requires morpho-physiological traits that are cheap and easy to score, characterized by a high genetic correlation with biomass production and by a heritability higher than that of productivity [77]. A common approach to determine the genetic basis of productivity and to identify efficient early predictors of biomass production is (1) to estimate the extent of genetic variability within different species; and (2) to study intra- as well as interspecific hybrids, with a view to determine how the combination of qualities of the parental species can give rise to hybrid vigor under particular environmental circumstances.

Consideration of genotype x environment interactions is important when formulating a forest tree breeding program, especially in the case of hybrid poplar, because poplar stands are generally propagated as clones, which have limited population buffering and thus limited capacity for homeostasis. Genotype x environment interactions imply that genotypes respond differently to different environments, so that the relationships between the effects of genotypes and environments are statistically not additive. Genotype x environment interactions may arise in two possible ways: (i) an array of genotypes rank differently in different environments; and (ii) the real difference between genotypes varies in magnitude between environments without changing the ranking. The latter type poses no problem because the best plants selected in one environment would still perform the best in another environment regardless of the differential expression of genotypes between environments. For poplar, genotypes for which yield remains high in spite of instability across environments are preferred to genotypes for which the yield is low and stable across environments. However, it is interesting to define environmental ranges, in which the ranking of the genotypes in terms of biomass production will remain stable. Most reports indicate that low heritability for any of the traits reflecting biomass production is explained by high genotype x environment interactions, and that most of the interactions occurs at the individual rather than at the family or provenance level [25, 48].

Within the POPYOMICs project, productivity and growth performance of the five pedigrees (*P. trichocarpa* '93-968' x *P. deltoides* 'ILL-129', POP1, 205 F_2 genotypes; *P. deltoides* '73028-62' x *P. trichocarpa* '101-74', POP2, 330 F_1 genotypes; *P. deltoides* 'S9-2' x *P. nigra* 'Ghoy', POP3a, 180 F_1 genotypes; *P. deltoides* 'S9-2' x *P. trichocarpa* 'V24', POP3b, 182 F_1 genotypes; *P. alba* '6K3' x *P. alba* '14P11', POP4, 142 F_1 genotypes) at the three sites (Headley, UK; Orléans, France; Cavallermaggiore, Italy) have been estimated both **directly**, through the determination of the biomass production at the end of the second growing season and the tree dimensions (stem height and circumference) at the end of each growing season, and **indirectly**, through the measurements of two categories of traits known to be closely related to biomass production: phenology (dates of bud set and bud burst) and canopy architecture (sylleptic ramification and resprouting ability after coppicing).

2.5 References

1. Cannell MGR (2003) Carbon sequestration and biomass energy offset: theoretical, potential and achievable capacities globally, in Europe and the UK. *Biomass and Bioenergy*, 24(2): 97-116

2. Tuskan GA (1998) Short rotation woody crop supply systems in the United States: What do we know and what do we need to know? *Biomass and Bioenergy*, 14, 307-315

3. El Bassam N (1998) Energy Plant Species: Their uses and impact on environment and development, James and James (Science Publishers) Ltd, London.

4. Sims REH, Maiava TG, Bullock BT (2001) Short rotation coppice tree species selection for woody biomass production in New Zealand. *Biomass and Bioenergy*, 20, 329-335

5. Heilman PE, Ekuan G, Fogle D (1994) Above- and below-ground biomass and fine roots of 4-year-old hybrids of *Populus trichocarpa* x *Populus deltoides* and parental species in short rotation culture. *Canadian Journal of Forest Research*, 24, 1186-1192

6. Rae AM, Robinson KM, Street NR, Taylor G (2004) Morphological and physiological traits influencing biomass productivity in short-rotation coppice poplar. *Canadian Journal of Forest Research,* 54, 335-345

7. Barigah TS, Saugier B, Mousseau M, Guittet J, Ceulemans R (1994) Photosynthesis, leaf area and productivity of 5 poplar clones during their establishment year. *Annals of Forest Science*, 51, 613-625

8. Bergez, JE, Auclair D, Bouvarel L (1989) First year growth of hybrid poplar shoots from cutting or coppice origin. *Forest Science*, 35, 1105-1113

9. Scarascia-Mugnozza GE (1991) Physiological and morphological determinants of yield in intensively cultured poplar (*Populus* spp.). Ph.D. thesis, University of Washington, Seattle, WA.

10. Scarascia-Mugnozza GE, Ceulemans R, Heilman PE, Isebrands JG, Settler RF, Hinckley TM (1997) Production physiology and morphology of *Populus* species and their hybrids grown under short rotation. II. Biomass components and harvest index of hybrid and parental species clones. *Canadian Journal of Forest Research*, 27, 285-294

11. Taylor G (2002) *Populus*: Arabidopsis for forestry. Do we need a model tree? Annals of Botany 90: 681-689. Tsarouhas V, Gullberg U, Lagercrantz U (2002) An AFLP and RFLP linkage map and quantitative trait locus (QTL) analysis of growth traits in Salix. *Theoretical and Applied Genetics,* 105, 277-288

12. Boerjan, W (2005) Biotechnology and the domestication of forest trees. *Current Opinions in Biotechnology*, 16, 159-166

13. Brunner AM, Busov VB Strauss SH (2004) Poplar genome sequence: functional genomics in an ecologically dominant plant species. *Trends in Plant Science*, 9, 49-56

14. Taylor G (2002) *Populus*: Arabidopsis for forestry. Do we need a model tree? *Annals of Botany,* 90, 681-689.

15. Wullschleger SD, Jansson S, Taylor G (2002) Genomics and forest biology: Populus emerges as the perennial favourite. *Plant Cell*, 14, 2651-2655

16. Andersson A, Keskitalo J, Sjodin A, Bhalerao R, Sterky F, Wissel K, Tandre K, Aspeborg H, Moyle R, Ohmiya Y, Bhalerao R, Brunner A, Gustafsson P, Karlsson J, Lundeberg J, Nilsson O, Sandberg G, Strauss S, Sundberg B, Uhlen M, Jansson S, Nilsson P (2004) A transcriptional timetable of autumn senescence. *Genome Biology*, 5, art no R24.

17. Updegraff K, Baughman MJ, Taff SJ (2004) Environmental benefits of cropland conversion to hybrid poplar: economic and policy considerations. *Biomass and Bioenergy*, 27, 411-428

18. Neale DB, Savolainen O (2004). Association genetics of complex traits in conifers. *Trends in Plant Science*, 9, 325-330

19. Madgwick HAI (1971) The accuracy and precision of estimates of the dry matter in stems, branches and foliage in an old-field *Pinus virginiana* stand. Miscellaneous Publication 132, Agricultural Experiment Station, University of Maine, Orono, 105-112.

20. Heilman PE, Xie F (1993) Influence of nitrogen growth and productivity of short-rotation *Populus trichocarpa* x *Populus deltoides* hybrids. *Canadian Journal of Forest Research*, 23, 1863-1869

21. Pontailler JY, Ceulemans R, Guittet J, Mau F (1997) Linear and non-linear functions of volume index to estimate woody biomass in high density young poplar stands. *Annales des Sciences Forestières*, 54, 335-345

22. Van Volkenburgh E, Taylor G (1996) Leaf growth physiology. In Biology of *Populus* and its implications for management and conservation ed. R.F. Stettler, H.D. Bradshaw, P.E. Heilman, T.M. Hinckley, NRC Research Press, Ottawa. 283-300.

23. Robinson KM, Karp A, Taylor G (2004) Defining leaf traits linked to yield in short-rotation coppice *Salix*. *Biomass and Bioenergy,* 26, 417–431

24. Scarascia-Mugnozza GE (1991) Physiological and morphological determinants of yield in intensively cultured poplar (*Populus* spp.). Ph.D. thesis, University of Washington, Seattle, WA.

25. Wu R, Bradshaw Jr HD, Stettler RF (1998). Developmental quantitative genetics of growth in *Populus*. *Theoretical and Applied Genetics*, 97, 1110-1119

26. Costes E, Guedon Y (2002) Modelling branching patterns on 1-year-old trunks of six apple cultivars. *Annals of Botany*, 89, 513-524

27. Newcombe G, Bradshaw HD, Chastagner GA, Stettler RF (1996) A major gene for resistance to *Melampsora medusae* f.sp. *deltoidae* in a hybrid poplar pedigree. *Phytopathology*, 86, 87-94

28. Jorge V, Dowkiw A, Faivre-Rampant P, Bastien C (2005) Genetic architecture of qualitative and quantitative *Melampsora larici-populina* leaf rust resistance in hybrid poplar: genetic mapping and QTL detection. *New Phytologist*, 167, 113-127

29. Wu R, Stettler RF (1998) Quantitative genetics of growth and development in *Populus*. III. Phenotypic plasticity of crown structure and function. *Heredity*, 81, 124-134

30. Wu R, Bradshaw Jr HD, Stettler RF (1998). Developmental quantitative genetics of growth in *Populus*. *Theoretical and Applied Genetics*, 97, 1110-1119

31. Ferris R, Long L, Bunn SM, Robinson KM, Bradshaw HD, Rae AM, Taylor G (2002) Leaf stomatal and epidermal cell development: identification of putative quantitative trait loci in relation to elevated carbon dioxide concentration in poplar. *Tree Physiology*, 22, 633-640

32. Rae AM, Ferris R, Tallis MJ, Taylor G (2006). Elucidating genomic regions determining enhanced leaf growth and delayed senescence in elevated CO_2. *Plant Cell and Environment*, 29, 1730-1741

33. Bradshaw HD, Villar M, Watson BD, Otto KG, Stewart S, Stettler RF (1994). Molecular genetics of growth and development in *Populus*. III. A genetic linkage map of a hybrid poplar composed of RFLP, STS and RAPD markers. *Theoretical and Applied Genetics*, 89, 167-178

34. Wullschleger SD, Yin TM, Difazio SP, Tschaplinski TJ, Gunter LE, Davis MF, Tuskan GA (2005). Phenotypic variation in growth and biomass distribution for two advanced-generation pedigrees of hybrid poplar. *Canadian Journal of Forest Research*, 35, 1779-1789

36. Frewen BE, Chen THH, Howe GT, Davis J, Rohde A, Boerjan W, Bradshaw Jr HD (2000). Quantiative trait loci and candidate gene mapping of bud set and bud flush in *Populus*. *Genetics,* 154, 837-845

37. Morreel K, Goeminne G, Storme V, Sterck L, Ralph J, Coppieters W, Breyne P, Steenackers M, Georges M, Messens E and Boerjan W (2006) Genetical metabolomics of flavonoid biosynthesis in Populus: a case study. *The Plant Journal,* 47, 224–237

38. Tsarouhas V, Gullberg U, Lagercrantz U (2004) Mapping of quantitative trait loci (QTLs) affecting autumn freezing resistance and phenology in Salix. *Theoretical and Applied Genetics,* 108, 1335-1342

39. Frewen BE, Chen THH, Howe GT, Davis J, Rohde A, Boerjan W, Bradshaw Jr HD (2000). Quantiative trait loci and candidate gene mapping of bud set and bud flush in *Populus*. *Genetics*, 154, 837-845

41. Jorge V, Dowkiw A, Faivre-Rampant P, Bastien C (2005) Genetic architecture of qualitative and quantitative *Melampsora larici-populina* leaf rust resistance in hybrid poplar: genetic mapping and QTL detection. *New Phytologist*, 167, 113-127

42. Zhang X, Wu N, Li C (2005) Physiological and growth responses of *Populus davidiana* ecotypes to different soil water contents. *Journal of Arid Environments*, 60, 567-5

43. Bungart R, Huttl RF (2004) Growth dynamics and biomass accumulation of 8-year-old hybrid poplar clones in a short-rotation plantation on a clayey-sandy mining substrate with respect to plant nutrition and water budget. *European Journal of Forest Research*, 123(2), 105-115

44. McCamant T, Black RA (2000) Cold hardiness in coastal, montane, and inland populations of *Populus trichocarpa Canadian Journal of Forest Research*, 30, 91-99

45. Crow P, Houston TJ (2004) The influence of soil and coppice cycle on the rooting habit of short rotation poplar and willow coppice. *Biomass and Bioenergy*, 26(6), 497-505

46. Li BL, Wu RL (1997) Heterosis and genotype x environment interactions of juvenile aspens in two contrasting sites. *Canadian Journal of Forest Research*, 27, 1525-1537

47. Ikuko N, Masahiko O (2001) Geographical transition of sylleptic/proleptic branching in three *Cinnamomum* species with different bud types. *Annals of Botany*, 87, 35-45

48. Lin CY, Togashi K (2002) Genetic improvement in the presence of genotype by environment interaction. *Animal Science Journal,* 73, 3-11

49. Basford KE, Cooper M (1998) Genotype x environment interactions and some considerations of their implications for wheat breeding in Australia. *Australian Journal of Agricultural Research*, 49(2), 153-174

50. Lin CS, Binns MR (1991). Genetic properties of four types of stability parameter. *Theor. Appl. Genet.*, 82, 505–509

51. Via S, Gomulkkiewicz R, de Jong G, Scheiner SM, Schlichting CD, van Tienderen PH, (1995). Adaptive phenotypic plasticity: consensus and controversy. *Trend Ecol. E*, 10, 212–217

52. Emebiri LC, Moody DB (2006). Heritable basis for some genotype-environment stability statistics: Inferences from QTL analysisof heading date in two-rowed barley. *Field Crops Research*, 96, 243-251

53. Kraakman ATW, Niks RE, Van den Berg PMMM, Stam P, van Eeuwijk FA (2004) Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. *Genetics*, 168, 435–446

54. Emebiri LC, Moody DB (2006). Heritable basis for some genotype-environment stability statistics: Inferences from QTL analysisof heading date in two-rowed barley. *Field Crops Research*, 96, 243-251

55. Papadakis JS (1984) Advances in the analysis of field experiments. *Proceedings of the Academy of Athens,* 59, 326-342

56. Henderson, C.R. (1953) Estimation of variance and co-variance components. *Biometrics,* 9, 226–252

57. M. Singh, S. Ceccarelli and J. Hamblin, (1993) Estimation of heritability from varietal trials data. *Theor. Appl. Genet.,* 86, 437–441

59. Via S, Conner J (1995) Evolution in heterogeneous environments - genetic-variability within and across different grains in tribolium-castaneum. *Heredity*, 74, 80-90 .

60. Mandel J (1969) The partitioning of interaction in analysis of variance. *Journal of Research of the National Bureau of Standards*, Section B - Mathematical Sciences, 73B, 309-328

61. Gauch HG (1992) Statistical analysis of regional yield trials: AMMI analysis of factorial designs. Elsevier, Amsterdam, The Netherlands.

62. Nachit MM, Nachit G, Ketata H, et al. (1992). Use of AMMI and linear regression models to analyze genotype-Environment interaction in Durum wheat. *Theoretical and Applied Genetics*, 83, 597-601.

63. Ebdon JS, Gauch HG Jr. (2002) Additive main effect and multiplicative interaction analysis of national turfgrass performance trials: I. Interpretation of genotype x environment

interaction. *Crop Science*, 42, 489-496

64. Remphey WR, Powell GR (1985). Crown architecture of Larix-Laricina saplings – sylleptic branching on the main stem. *Canadian Journal of Botany*, 63(7), 1296-1302

65. Nyquist WE (1991). Estimation of heretability and prediction of selection response in plant-populations. *Critical Reviews in Plant Science*, 10, 235-322

66. Zabek LM, Prescott CE (2006). Biomass equations and carbon content of aboveground leafless biomass of hybrid poplar in Coastal British Columbia. *Forest Ecology and Management,* 223, 291-302

67. Eckenwalder JE, Heath MC (2001) The evolutionary significance of variation in infection behavior in two species of rust fungi on their hosts and related nonhost plant species. *Canadian Journal of Botany*, 79, 570-577

68. Schulte PJ, Hinckley TM, Stettler RF (1987). Stomatal responses of *Populus* to leaf water potential. *Canadian Journal of Botany*, 65, 255-260

69. Cited in: Wu, R, Stettler, RF (1997) Quantitative genetics of growth and development in Populus. II. The partitioning of genotype x environment interaction in stem growth. *Heredity,* 78, 124-134

70. Cited in Halle F (1986). Modular growth in plants. *Philosophical transactions of the Royal Society*, 313, 77-87

72. Rohde A, Bhalerao RP (2006) Plant dormancy in the perennial context. *Trends in Plant Science,* 12, 217-223

73 . Wricke G (1962) Über eine Methode zur Erfassung der ökologischen Streubreite in Feldversuchen. *Zeitschur f Pflanzenz*, 47, 92–96

74. Rae, A.M., Pinel, M.P.C., Bastien, C., Sabatti, M., Street, N., Tucker, J., Dixon, C., Marron, N., Dillen, S.Y., Taylor, G., (2007) QTL for yield in bioenergy *Populus*:

identifying G × E interactions from growth at three contrasting sites. Tree Gene. Gen., doi:10.1007/s11295-007-0091-3.

75. Marron **N**, Bastien C, Sabatti M, et al. (2006) Plasticity of growth and sylleptic branchiness in two poplar families grown at three sites across Europe. *Tree Physiology*, 26(7), 935-946

76. Marron **N**, Dillen SY, Ceulemans R (2007). Evaluation of leaf traits for indirect selection of high yielding poplar hybrids. *Environmental and Experimental Botany*, 61, 103-116

77. Tuberosa R, Salvi S, Sanguineti MC, et al. (2002) Mapping QTLs regulating morpho-physiological traits and yield: Case studies, shortcomings and perspectives in drought-stressed maize. *Annals of Botany*, 89, 941-963

3. QTL discovery for yield, disease and pest traits in *Populus*

V Jorge, V Storme, AM Rae, M Gaudet, I Paolucci, C Bastien, G Taylor, M Sabatti

3.1 Introduction

Fast-growing *Populus* hybrids are being developed as biofuels owing to their rapid juvenile growth [1], high photosynthetic capacity [2], superior growth performance [3] and large woody biomass production [4].

The woody genus *Populus* offers extremely important potential not only as a commercially grown tree, but it is also of tremendous value as the model tree, due to its relatively small genome size (400-500Mb), with a highly developed number of molecular genetic maps, combined with the ability of easy genetic transformation. *Populus* may be propagated vegetatively, making mapping populations immortal and easing the ability to produce large amounts of clonal material for experiments. [5]. Hybridisation occurs routinely and in these respects it has many similarities to Arabidopsis. The development of large EST collections and microarray analysis, the best available globally for any tree and the availability of mapping pedigrees for QTL detection secure *Populus* as the ideal subject for further exploitation. Of crucial importance is the availability of the poplar genome sequence released in 2004 and detailed further in Chapter 5.

A significant problem is the *consistent* attainment of high yields that are commercially viable in large scale planting, across a wide variety of climatic and site conditions. Despite trial yields of 35 ODT ha^{-1} y^{-1} [6], and theoretical potential being calculated between 40 and 50 ODT ha^{-1} y^{-1} [7], many commercially grown plots yield between 10 and 15 ODT ha^{-1} y^{-1}. Long-term and widespread planting of energy crops, therefore, requires a supply of genetically improved material, suitable for growth at a large number of temperate sites.
Fundamental to the breeding of poplar as a bioenergy crop is the understanding of the genetic control of traits influencing biomass. Traits such as height and stem diameter can be used to non destructively estimate biomass yield, but many other traits are involved [8]. Overall biomass gain is a function of many internal plant processes and their interactions with the environment. It is, therefore, important to study the relationships and interactions between

biomass and other important correlated traits. These traits are usually complex requiring knowledge of quantitative genetics to resolve their action.

The difficulties of QTL discovery in trees can never be underestimated – they are outbreeding and dioecious. The quantitative genetics of outbreeding species is complicated by their heterozygousity so that up to four alleles may be segregating at each locus. Quantitative Trait Loci (QTL) analysis can be used to understand the genetic control of complex traits, but QTL mapping experiments usually give poor resolution of position, and verification of the presence of QTL is rare making further study difficult. Increased competence of QTL mapping has been explored by using methods such as interval mapping [9, 10], replicate testing by repeating experiments over several years, repeating across different environments, and comparison across different pedigrees. It is possible that many traits measured for QTL mapping are related. For example, it is expected that a plant which shows certain behaviour for height at the first measurement would show similar behaviour at a latter stage of development or that leaf length is related to leaf width. Correlations between traits may be due to genetic linkage, to allometry of development or to pleiotropy. Selective breeding requires that advantageous characteristics can be selected without the introduction of disadvantageous traits, therefore it is necessary to know whether correlated characters can be separated by selective breeding. Another reason for the study of correlation between traits is that a common QTL influencing multiple phenotypic measures is generally easier to detect in multivariate than in univariate data, because adding another measured indicator of the latent trait to the model can, in some circumstances, have the same effect as doubling the sample size. Multicollinearity exists when one or more variables can be predicted with high precision from one or more of the remaining independent variables, giving a redundancy of information, and complicating the manipulation of the independent variables.

Melampsora rust is considered as the most widely distributed and serious foliar disease of the Aigeiros and Tacamahaca poplars and their hybrids, causing premature defoliation and shoot growth reduction. Rust epidemics, in combination with secondary parasites such as *Discosporium populeum* and *Cytospora chrysosperma*, can cause severe economic loss and even plant death. In Europe, selection of poplar clones showing complete (qualitative) resistance to *M.larici-populina* combined with the use of these clones in monoclonal plantations, let to the regularly emerging of new races that have overcome the resistance. Recently, this occurred on a large scale with the clone ‘Beaupré’, a cultivar constituting 80%

of the poplar plantations in Belgium and the North of France. *Xanthomonas populi,* the causal agent of poplar stem and branches oozing canker disease the other major threat for the cultivation of poplar in North -Western Europe . These pests and diseases must also be understood at a genetic level to enable improved breeding and selection of superior material for future planting.

The objective of this work was to test the robustness of QTL in contrasting genetic backgrounds and environments. Our approach was to use the new experimental network with identical sites in the UK, France and Italy, as described in Chapter 1, to identify QTL for yield traits for disease and pest resistance, in contrasting genetic backgrounds. Previous studies on QTL in *Populus* have generally been done in one pedigree and most often in one environment [11-17]. These have been useful, but here we wish to extend this analysis to identify robust QTL that occur over diverse genetic backgrounds and also to indentify QTL that may only be present in certain environments. For example, in a recent study within the consortium, QTL in response.

We present here an overview of the results of the QTL analysis performed using 5 different pedigrees, 10 maps, 11 traits for physiology and yield, and 22 traits for disease and pest resistance and 3 different environments. This dataset has considerable depth and will be of value in identifying regions of the *Populus* genome for future breeding and improvement.

3.2 MATERIAL AND METHODS

Genetic map constructions

POP1. Genotypic segregation data from simple sequence repeat (SSR) and amplified fragment length polymorphism (AFLP) markers were used to construct a sex-average framework map for the F_2 Family pop1 (Sewell et al. unpublished data). Ninety one SSRs genotyped on 350 individuals and 92 fully informative AFLPs genotyped on 165 individuals were selected from a previous mapping study of Family 331 [18]. Fully informative markers (i.e., a marker that is heterozygous for different alleles in each of the F_1 parents) were preferentially chosen when available. The framework map was constructed at LOD 5 using MapMaker [19] and JoinMap [20]. Methods pertaining to SSR and AFLP analyses for *Populus* followed [18] and [22], respectively. Map construction followed [23]. The total map distance was 1453.1 cM with an average of 8 cM between markers. Linkage disequilibrium was seen to occur over relatively small regions. For QTL mapping using multiQTL the sex

average map was converted to two parental pseudo testcross maps [24]. Two hundred and ten individuals were used for QTL mapping.

POP2. Three hundred and thirty six individuals of the full-sib pop2 family (*P. deltoides* 73-028-62 x *P. trichocarpa* 101-74) were genotyped with RFLP, AFLP, RAPD, STS and SSR as described in Jorge et al. (2005) and in the material and methods section of 'genomic technologies'. After data scoring, segregation distortion was checked with a χ^2 test and highly significantly ($P<0.01$) skewed markers were discarded. A genetic marker map was constructed with the software MAPMAKER 3.0b [19] using the pseudo-testcross strategy described by Grattapaglia and Sederoff [24]. The marker dataset for each parental map was duplicated to allow the detection of linkage for markers in the repulsion phase. Analyses were performed with a LOD score threshold of 4.0 and a maximum of recombination value θ = 0.40 for grouping using the "group" command. For the framework map, the best marker order of the linkage group having eight or fewer markers was identified using the "compare" command, whereas the order of the groups with more than eight markers was identified using the "order" and "try" commands. The marker order of each linkage group was verified using the "ripple" command. In all cases, a difference of -2.0 for the likelihood was used as threshold. Kosambi's mapping function was applied for map distance calculation. For the complete genetic map, accessory markers were mapped on the framework map according to the most likely position.

POP3a and POP3b. The interspecific crosses between *P. deltoides* cv. S9-2 and *P. nigra* cv. Ghoy (=POP3a) and between *P. deltoides* cv. S9-2 and *P. trichocarpa* cv. V24 (=POP3b) were repeated in 1995 and 1997 to provide more offspring. The resulted full-sib families were named 95001, 97001, 95002 and 97002. The first 2 digits refer to the year of crossing, '001' refers to pop3a, '002' refers to pop3b. AFLP analysis was performed as described in [25] on an extra 96 individuals for POP3a and an extra 114 individuals for pop3b. A χ^2 test (d.f. =1, $P<0.01$) was used to identify deviations from Mendelian ratios. AFLP markers deviating at the 1% significance level were excluded from the linkage analysis. Linkage analysis was performed with MAPMAKER Unix version 3.0 [19] as described before [25]. Markers ordered at a LOD score of 2.0 were used as framework markers. Markers that could not be ordered with equal confidence were positioned relatively to a framework marker and are

called accessory markers. Microsatellites are used as bridge markers to align the genetic map to the sequence map of *Populus* and to the other genetic maps.

POP4. Two genetic linkage maps were constructed for the species *Populus alba* using an intra-specific F1 population. 141 F1 individuals were genotyped with AFLP, RAPD and SSR markers. Two linkage maps (one for each parent) were constructed using double pseudo-testcross strategy [24] and software Mapmaker 3.0 (Lander et al., 1987)[19]. Markers of each parental map were grouped at LOD 3.5 and maximum distance 37.2. "Compare", "order" and "ripple" commands were used with a LOD-threshold of 2 to find the correct order of framework markers in the linkage groups. Accessory markers were added to the framework markers using the "try" command and placed in the most probable position. SSR markers were used as bridge markers to align the genetic maps to the sequence map of *Populus* and to the other genetic maps.

POP5. 92 individuals of the full-sib pop5 family (*P. nigra* 58-861 x *P. nigra* Poli) were genotyped with 533 AFLP markers, 130 SSRs, and 7 SNPs. For each marker a χ^2 test (d.f. = 1, $P < 0.01$ and $P < 0.05$) was performed to identify deviation from Mendelian ratios. AFLP markers deviating at 1% significance level were excluded from the linkage analysis. The other markers (deviating at $0.01 < P < 0.05$) were noted as distorted but conserved in the data set. The two-way pseudo-testcross mapping strategy [26, 24] was performed with MapMaker version 3.0 [19]. To detect linkages in repulsion phase, the marker data set for each parent was inverted and added to the original data. The "triple error detection" and the "error detection" features were used. Markers were grouped (group command) and ordered (order command) using a LOD score threshold of 4.0 and a maximum recombination fraction θ of 0.3. For linkage groups with more than 5 markers, the "three point" command was used to pre-compute the likelihood of all three point crosses of each group. A framework map was established with the 'order' command' when marker order was supported by a log-likelihood ratio support of 2.0. The order of the markers was controlled with the "ripple" command. Additional markers were added with the "try" command. For the linkage groups with less than 5 markers, the "compare" command was used. For the complete genetic map, markers that could not be ordered with equal confidence and markers distorted at $P < 0.01$ were indicated as accessory and mapped on the framework map according to the most likely position. Distances between marker loci were calculated from recombination fraction using Kosambi's mapping function.

The number of framework markers, accessory markers, the observed (framework) map length, and the average distance between two (framework) markers are listed in Table 3.1. Only linkage groups with 3 or more markers (frame and accessory) were considered. Maps are visualised with the web-based tool CMAP (http://www.gmod.org) and can be seen on https://www.appliedgenomics.org/cmap/popyomics.html.

Table 3.1. Summary table for the complete genetic maps and for the framework maps of all populations. 'LG' Linkage groups, 'T' *P. trichocarpa* , 'D' *P. deltoides*, 'N' *P. nigra*, 'A' *P. alba.*

	POP1			POP2		POP3a		POP3b		POP4		Pop5	
	F2 family 331	TxD ♀	TxD ♂	D ♀	T ♂	D ♀	N ♂	D ♀	T ♂	A ♀	A ♂	N ♀	N ♂
Total number of markers	182	119	135	363	306	459	367	363	342	299	297	368	317
Total number of framework markers	182	119	135	230	238	213	225	166	197	203	219	195	188
Number of LG including all markers	22	20	21	33	35	21	28	23	23	24	30	20	25
Number of LG including only framework markers	22	20	21	29	39	21	28	23	23	23	29	18	25
Observed map length (cM)	1478.9	3018.9	2846.4	4051.8	3661.1	2183.2	2877.9	1863.0	2183.8	3006	3447	2104.1	2467.3
Observed framework map length (cM)	1478.9	3018.9	2846.4	3209.8	3282.8	2112.8	2609.9	1758.3	2114.9	2599	2892	1917	2195.4
Mean interval distance and standard deviation (cM)	9.2±9.0	30.5±42.5	29.5±52.5	12.3±8.3	13.5±8.7	5.0±6.4	8.6±8.9	5.5±7.0	6.8±7.3	12.4±8.5	13.3±9.7	6.0±6.5	8.4±8.5
Mean interval distance between framework markers and standard deviation (cM)	9.2±9.0	30.5±42.5	29.5±52.5	16.0±8.7	16.5±9.7	11.0±7.7	13.7±8.0	12.3±8.0	12.2±7.6	14.9±7.1	15.7±8.3	11.0±6.2	13.5±7.3

QTL analysis strategy

Comparison of QTL software programs

During a workshop on QTL analysis with partners P1, P2, P3 and P6, a comparison was made between four QTL programs: the commercial package MapQTL (http://www.kyazma.nl/, [27] , the free web-based program QTL Express (http://qtl.cap.ed.ac.uk/), the free downloadable program QTL cartographer (http://statgen.ncsu.edu/qtlcart/) , and the commercial package MultiQTL (http://www.multiqtl.com/, MultiQTL Ltd, Institute of Evolution, Haifa university, 31905 Haifa, Israel) . Finally the program MultiQTL was chosen for its ability to do QTL analysis across multiple environments, to test for two-linked QTL models, and to calculate confidence intervals. Simultaneous treatment of data from multiple environments provides a significant increase in power of QTL detection and accuracy of the estimated QTL position and effect [28]. This program was used for QTL analysis on the populations POP1, POP2, POP3a, POP3b, and POP4.

Trait selection

In total, 13 traits were measured on all populations at all sites (F, UK, I), 6 traits were measured on all populations at two sites (F, I), and 3 budset parameters were available for all populations at the French site, for pop4 also data was available at the Italian site. An overview of the traits is given in Table 3.2. The traits as well as the design of the plantations are described elsewhere. In addition, two stability parameters were calculated for each trait according to the additive and multiplicative model interaction (AMMI) (described elsewhere). Since a lot of these traits were highly correlated with each other, a selected number of traits was chosen for QTL analysis according to heritabilities values: Budburst, Height2, Circum2, Date15_CNL, Syllep1 and the corresponding stability parameters for Budburst, Circum2, and Syllep1.

Table 3.2 Overview of the yield traits analysed. [1]data for Italy is only available for POP4

abbreviation	definition	sites
Budburst	Budburst (day of year)	F,I,UK
Date15_cnl	date that score 1.5 is observed for first time, transition to bud structure	F, I[1]
Duration_subproc1_cnl	duration from the onset of bud setting process until the transition to bud structure	F, I[1]
Duration_subproc2_cnl	duration from the transition to bud structure until closed bud	F, I[1]
Height1	Stem height - Year 1 (cm)	F,I,UK
Circum1	Stem circumference - Year 1 (mm)	F,I,UK
Ht_Cc1	Ratio Height to Circumference (cm mm^{-1}) - Year 1	F,I,UK
Volume1	Volume of the stem assuming a cone - year 1 (cm^3)	F,I,UK
Height2	Height - Year 2 (cm)	F,I,UK
Circum2	Circumference - Year 2 (mm)	F,I,UK
Ht_Cc2	Ratio Height to Circumference - Year 2 (cm mm^{-1})	F,I,UK
Volume2	Volume of the stem assuming a cone - year 2 (cm^3)	F,I,UK
DeltaCircum	Increment in circumference between the two growing seasons (mm)	F,I,UK
DeltaHeight	Increment in height between the two growing seasons (cm)	F,I,UK
Syllep1	Number of sylleptic branches - Year 1	F,I,UK
Syltotop	Distance to top of the sylleptic branches - Year 1 (cm)	F,I,UK
ABPrealDW	Realised aboveground biomass production (ton ha^{-1} $year^{-1}$) - Dry weight basis	F,I
ABPrealDWstem	Realised aboveground stem dry weight biomass production (ton ha^{-1} $year^{-1}$)	F,I
TreeDW	Tree aboveground dry weight (gram $year^{-1}$)	F,I
StemDW	Stem dry weight (gram $year^{-1}$)	F,I
Stem_Total	Ratio Stem / Total aboveground dry weight biomass	F,I
Shoots	Number of resprouts	F,I

Additionally, we present here QTL results for the disease and pest resistance traits showed in Table 3. Results on comparative mapping focused on field rust resistance (leaf score, **LS**) on POP1, POP2, POP3a and POP3b; (see previous reports for detailed protocols).

Table 3.3 Overview of the disease and pest traits analysed. [a] the 5 strains are 98AG69, 98AR1, 93ID6, 93CV1, 93JE3, inoculated as described in [30]. [b] Evaluation of LS in Italy only for POP2 and POP3a and b.

Disease or pest	Abbreviation	Definition	Sites(/strains)	Population
Rust	LP	latent period	Laboratory/5 strains[a]	POP2
Rust	UN	Number of uredinia	Laboratory/5 strains[a]	POP2
Rust	US	size of uredinia	Laboratory/5 strains[a]	POP2
Rust	LS	leaf score	F, I[b], UK	POP1, POP2, POP3a, POP3b
Rust	Def	rust defoliation	F	POP2
Rust	Glo	global score	F	POP2
Rust	TolCir3	Tolerance in circumference	F	POP2
Rust	TolH5	Tolerance in height	F	POP2
Rust	TolVol	Tolerance in volume	F	POP2

Canker	GI	girdling index	F	POP2
Canker	LL	Lesion length	F	POP2
Chrysomela	chryso	log of leaf consumed area	Laboratory	POP2

QTL analysis

Single trait analysis was performed using first the interval mapping approach followed by multiple interval mapping. The combination of MIM with the multiple environment approach increases the accuracy of the estimated QTL position. The problem for the MIM approach is how to choose QTL to fit into the MIM model. The selection procedure here was to scan the entire genome for one-linked QTLs and then for two-linked QTLs. All QTLs with a significant p-value < 0.05 were included in the model. The option 'marker restoration' was used to reduce the effect of missing information. MultiQTL recalculates the maps based on genotypic data. The Kosambi mapping function was chosen. Because of lack of an 'error detection' option, as provided in Mapmaker, the maps calculated with MultiQTL are expanded compared to the maps calculated by Mapmaker. The strategy, as outlined in Fig. 3.1, starts with scanning the entire genome for QTLs assuming a single model i.e. one QTL per linkage group. Permutation tests (1000 runs), comparing hypotheses H1 (there is a QTL in the chromosome), and H0 (no QTL in the chromosome) were run to obtain chromosome-wise statistical significance. In a second step the genome was scanned for QTLs assuming a two-linked QTL model. For chromosomes where a single QTL was already detected, permutation tests (1000 runs) were run to compare the hypotheses H2 (two linked QTLs in the chromosome) versus H1. Subsequently, when p(H2 vs H1)<0.05, permutations were run to compare H2 vs H0. If both hypothesis tests were significant, it was necessary to check whether the two detected intervals were in adjacent intervals. Such a result is most likely a false positive result (Korol, personal communication). A two-linked QTL model was only accepted if the two intervals were coinciding with the peaks of the single QTL model. For other chromosomes, the hypothesis H2 vs H0 was tested first. To speed up calculation time, permutations for the two-linked QTL models were only conducted with 1000 runs when the p-value was <0.1 after 100 runs. In a last step MIM was performed. For the remaining significant QTLs again all necessary permutations were run and bootstrap analysis was

performed to estimate the 95% confidence interval. The same strategy could be used for single site analysis. The framework maps of the 5 populations, as calculated with MultiQTL, were uploaded into CMAP together with the positions of the detected QTLs and the % of variance explained by the QTL.

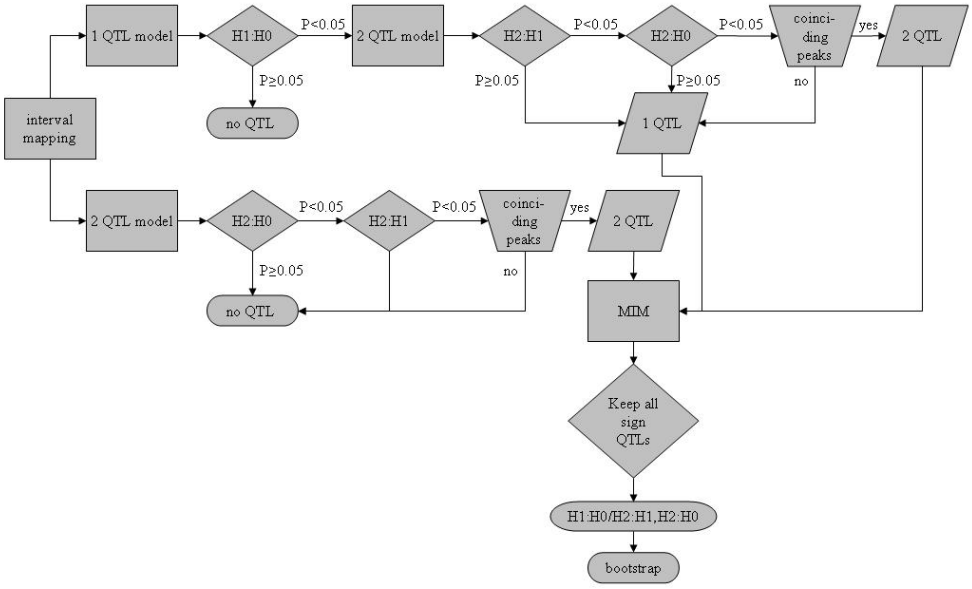


Fig. 3.1 Flowchart of the followed qtl strategy in MultiQTL. H1:H0 tests the hypothsesis of having 1 qtl in the chromosome compared to no qtls; H2:H0 tests the hypothesis of having 2 qtls in the chromosome compared to no qtls; H2:H1 tests the hypothesis of having 2 qtls in the chromosome compared to 1 qtl. The significance of the hypothesis tests were based on permutation tests. To minimise computational time, 1000 runs were only done when p<0.1 after 100 runs. More explanation is given in the text.

3.3 Results and Discussion

We report here QTL results for POP1, POP2, POP3a, POP3b and POP4, as POP5 was an extra pedigree not included in the contract.

QTLs for physiological and yield traits

QTLs for each pedigree

Table 3.4 summarizes QTL results for each pedigree for the selected number of trait chosen according to heritabilities values: Budburst, Height2, Circum2, Date15_CNL, Syllep1 and the corresponding stability parameters (MUL1 and MUL2) for Budburst, Circum2, and Syllep1 (see also Material and Methods).

Table 3.4 Summary of QTL results obtained in each of the 5 pedigrees studied. [a] Top number is the number of QTL detected and bottom numbers are the range of PEV; [b] Percentage of explained variance. NA: trait data not available.

	Pop 1		Pop2		Pop3a		Pop3b		Pop4		
	TxD ♀	TxD ♂	D ♀	T ♂	D ♀	N ♂	D ♀	T ♂	A ♀	A ♂	Total (PEV[b] range)
Budburst	7[a] 3-13	5 2-35	9 1.9-9.1	13 1-7.5	10 3.7-14.4	3 4.8-7.9	7 3.1-11.2	1 3.6	5 2.8-7.8	6 2.2-1.8	66 (1-35%)
Budburst MUL1	1 7	1 41	4 2.8-3.9	3 1.4-5.8	3 9.4-11.7	1 8.4	3 5.8-8	2 8.8-13.9	2 9.6-12.8	3 5.1-18.6	23 (1.4-41%)
Budburst MUL2	- -	1 8	2 3.2-10.4	2 2.1-2.4	1 12.9	2 10.2-12.7	1 4.4	2 6.1-9.2	2 9.8-12.1	1 13	14 (2.1-13%)
Height2	7 3-14	7 7-21	4 3.7-6.5	- -	2 7.9-13.6	5 6.4-6.8	5 3.2-6.7	2 7.3-25.4	4 2.2-13.1	4 5.9-8.9	40 (2.2-25.4%)
Circum2	7 3-22	3 7-14	7 1.1-3.4	8 3.3-10.1	5 5.3-18.1	7 4.2-14.9	7 4-8.4	2 5.6-18.4	4 5.6-17.6	2 7.3-15.3	52 (1.1-22%)
Circum2 MUL1	2 8-37	1 6	2 2.1-4.2	4 2.9-7	2 9.8-10.3	2 7.6-9.3	- -	1 6	1 15.9	1 10.3	16 (2.1-37%)
Circum2 MUL2	1 6	1 34	2 3.9-4.8	1 9.4	- -	4 5.4-12.8	2 6.2-8.7	1 16.9	1 17.1	1 15.7	14 (3.9-34)
Date15_CNL	NA	NA	4 3.5-9.5	3 2.2-31.3	1 17.1	2 6.9-7.7	3 6.6-7.6	2 7.5-8.6	2 4.8-7.7	3 4.8-18.8	18 (2.2-31.3%)
Syllep1	6 4-13	8 4-15	6 1.8-7.9	6 1.6-15.6	4 6.1-10	5 5.7-17.2	3 4.3-23	5 3.3-10.8	3 3.6-9.7	4 2.8-6.3	50 (1.6-23%)
Syllep1 MUL1	1 6	3 6-15	2 5.3-7.4	4 3-6.2	- -	2 7.8-9.3	3 6.2-7.4	1 6.8	3 9.1-10	- -	19 (3-15%)
Syllep1 MUL2	2 9-25	1 25	1 5.3	3 2.4-8.3	1 8.4	2 6.9-10.3	1 14	1 10.4	- -	2 9.1-14.6	12 (2.4-25%)

The total numbers of QTL for each trait (stability parameters excluded) varies from 18 to 66 with a range of PEV from 1% to 35%). Most significant QTLs were detected for budburst (up to 35% of explained variance) as expected because of high level of variation and heritability for this trait (see previous chapters) and also based on previous studies: for example, bud burst in Scots pine and poplar was determined by a few loci with major effects [27, 28, 13]. Most significant QTLs are detected also for POP1 pedigree (PEV of 15 to 35%), this could be due to the level of variation in this particular F2 segregating pedigree.

The AMMI genotypic scores (MUL1 and MUL2) reflect, for each genotype, their adaptation to specific environments. Any genotypes with values close to zero shows general adaptation to the tested environments, whereas large genotypic value reflects adaptation to a specific environment with score of the same sign. QTL mapping using these parameters could identify those regions of the genome responsible for differential

genotypic expression across environment. Results showed that some genome regions are explaining a large part of the variation (up to 41%) of these scores, for the three traits (budburst, circumference and number of sylleptics), particularly in POP1 and POP4 pedigrees. Similar results were found by [29] in barley.

Comparison of QTLs detected across pedigrees.
This comparison has been made for Budburst, Circum2, Date15_CNL, Syllep1 (Table 3.5).

Table 3.5 Numbers of QTL showing co-localisation (i.e. same linkage group) on 2 to 6 maps for 4 traits associated with yield (names of the linkage groups carrying the QTLs are indicated in parenthesis).

	2 maps	**3 maps**	**4 maps**	**5 maps**	**6 maps**
Budburst	4 (LGXIII, LGXVI, LGXVIII, LGXIX)	5 (LGI, LGIII, LGIV, LGIX, LGX)	3 (LGVI, LGXII, LGXIV)	4 (LGII, LGV, LGVII, LGVIII)	1 (LGXI)
Circum2	5 (LGIII, LGV, LGXI, LGXIII, LGXV)	3 (LGIV, LGXII, LGXVI)	4 (LGVII, LGX, LGXVII, LGXIX)	1 (LGXIV)	
Date15_CNL	3 (LGII, LGV, LGIX)	2 (LGXI, LGXII)			
Syllep1	7 (LGI, LGVI, LGVII, LGX, LGXIII, LGXIV, LGXVII)	1 (LGXV)	3 (LGIV, LGV, LGXVI)		1 (LGVIII)

We observed an important number of co-localisations but few were highly significant (i.e. co-localisation on 5 to 6 maps, out of 11 maps). We have also to underline that the number of co-localised increase with the total number of QTL detected (Budburst: 66; Circum2: 52; Date15_CNL: 18; Syllep1: 50); Thus, co-localisation could be simply the consequence of a random distribution of QTLs on the maps, and could not have functional significance.

Given these results of co-localisation of QTLs, we choose to focus future studies on QTLs that co-localised on 5 or 6 maps. This is the case for 3 traits for which QTLs explain low to moderate percentage of variance:

- For Budburst on LGXI, QTLs explain between 2.2% and 9.9% of the variation.
- For Circum2 on LGXIV, QTLs explain between 2.8% and 8% of the variation.

- For Syllep1 on LGVIII, QTLs explain between 4% and 17.2% of the variation.

The apparent robustness of these QTLs (shown on Figure 3.2) opens perspectives of a genetical genomic study (see Chapter 5), with the objective of isolating genes controlling these traits using map, genome and expressional information.

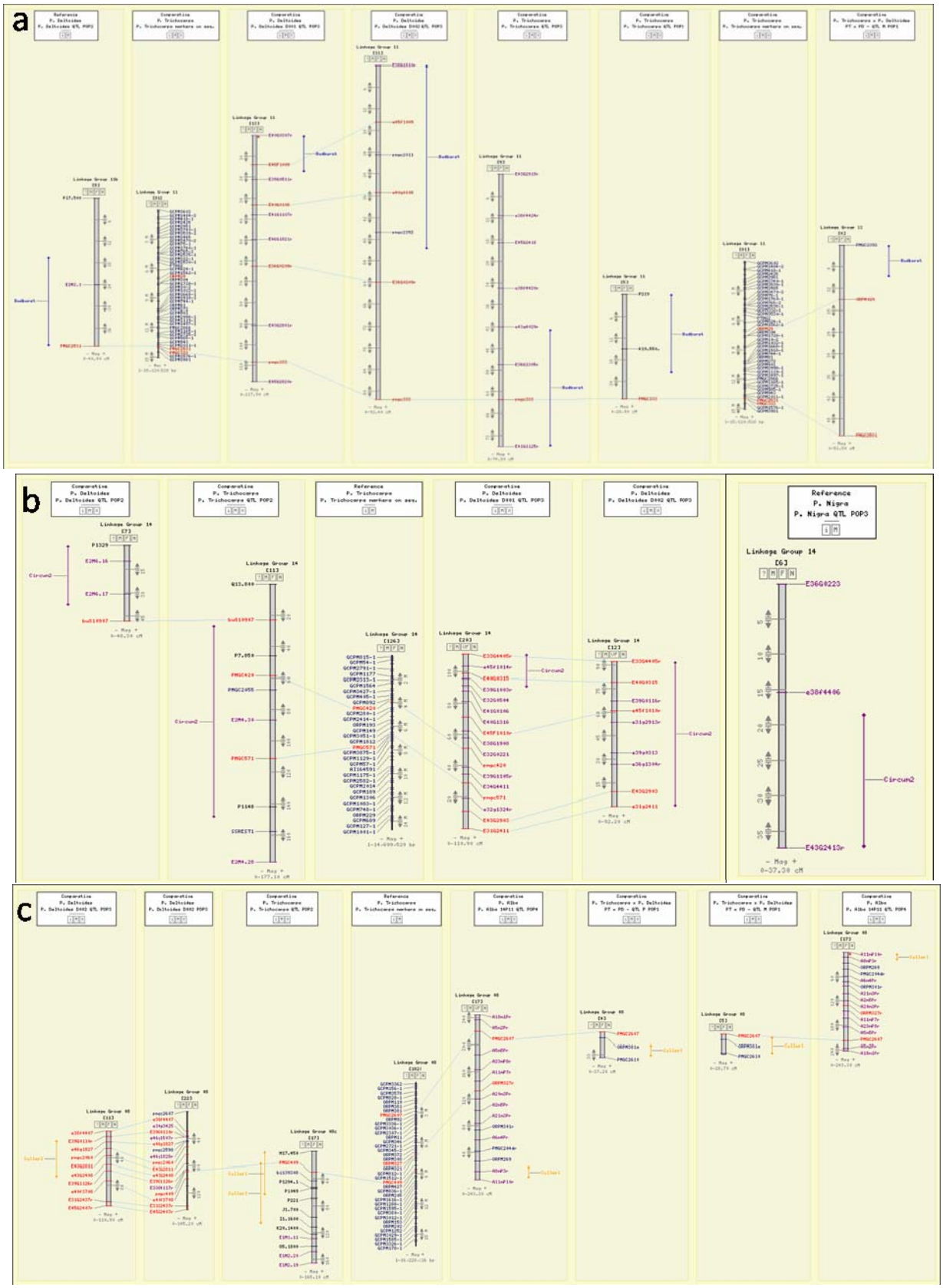

a
b
c
Reference
P. Nigra
P. Nigra QTL POP3
Linkage Group 14
E36G8223
e38f4486
Circum2
E43G2413r

Figure 3.2 Most significant co-localisations of QTLs for physiology and yield. Example of alignment and co-localisation of QTLs for a: Budburst; b: Circumference; c: Number of sylleptics, for different maps and pedigrees, and projection to genome sequence. Vertical black (genome sequence) and grey (genetic map) bars are linkage groups with markers positions; Vertical blue lines are confidence intervals of QTLs. For trait names in blue refer to previous chapters and materiel and methods. Correspondences between genetic maps and genome sequence are drawn as light blue lines.

QTL for disease and pest resistance traits

QTLs for several disease and pest for one pedigree (POP2).

QTLs detected for different components of quantitative resistance and tolerance to rust, to chrysomela and canker represent a total of 80 QTLs (Table 3.6).

Table 3.6 Summary of QTL results detected in POP2 for different components of quantitative resistance and tolerance to rust, to chrysomela and to canker (D: *P. deltoides* parental map; T: *P. trichocarpa* parental map).

Trait	Component	Number of QTL	Range of PEV	Localisation of the highest PEV value QTL
Rust	3 lab. components & strains	D: 16 T: 27	1.8 – 71.9 1.3 – 79.5	LGXIX LGXIX
Rust	3 components of tolerance	D: 6 T: 7	2.2 - 25 2.1 - 15.8	LGXVI LGI
Rust	Field resistance	D: 4 T: 4	1 - 76.4 1 – 72	LGXIX LGXIX
Chrysomela		D: 4 T: 1	3.5 - 24.4 8	LGXVI p (not aligned)
Canker	2 components	D: 4 T: 7	2.9 - 9.7 3.8 - 9.4	LGVII b (not aligned)

Data for rust resistance evaluated in laboratory has been partially published in [16]. Most significant QTLs on LGXIX are shown in Figure 3.3.

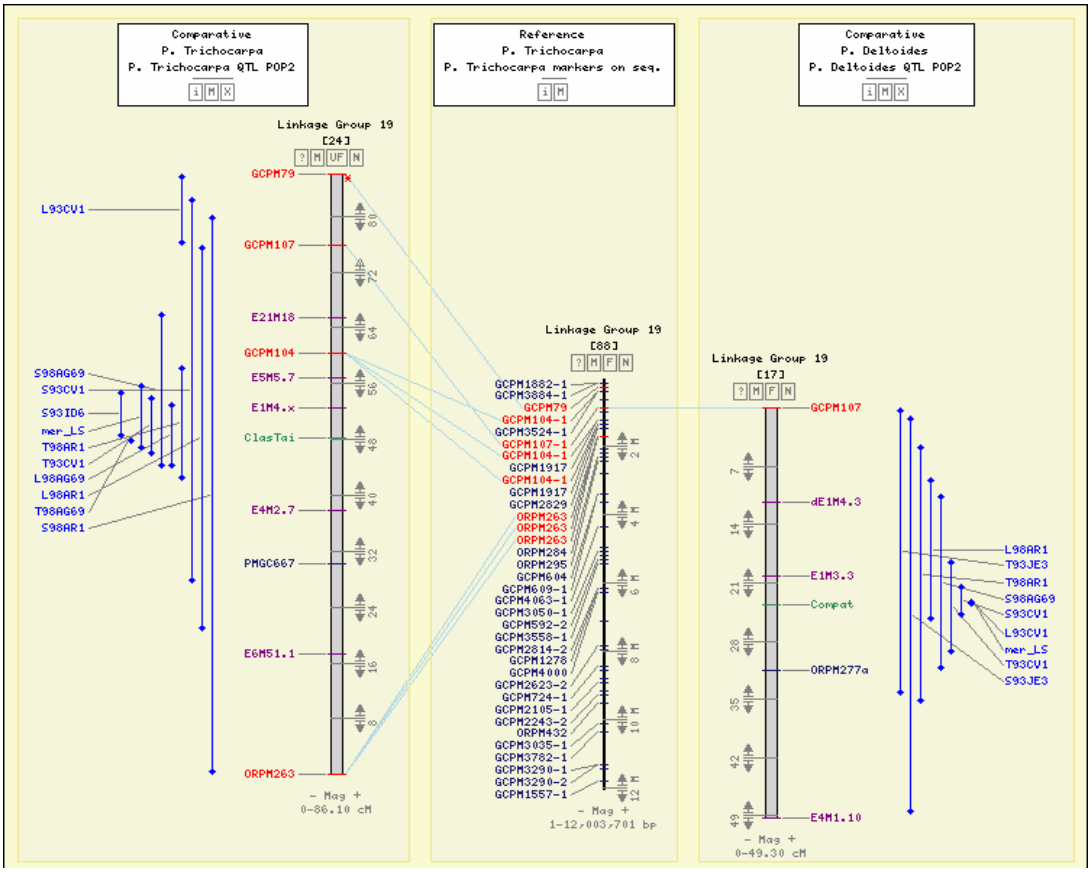


Figure 3.3 Most significant QTLs for rust resistance located on LGXIX in *P. deltoides* map (right map) and *P. trichocarpa* map (left map) in POP2 pedigree, and projection to genome sequence (map at the centre). Vertical black and grey bars are linkage groups with markers positions; Vertical blue lines are confidence intervals of QTLs. For trait names in blue refer to previous chapters. Correspondences between genetic maps and genome sequence are drawn as light blue lines.

We have analysed QTLs for different disease and pest resistance to determine whether they are co-localised in the same genome regions. Five linkage groups carry QTLs controlling resistance to 2 or 3 disease and pest resistance:

- LGII: Rust resistance (lab.), rust tolerance, canker.
- LGIV: Rust resistance (lab.), chrysomela, canker.
- LGVIII: Rust resistance (lab., field), chrysomela, canker.
- LGXII: Rust resistance (lab., field), rust tolerance, chrysomela, canker.
- LGXVI: Rust resistance (lab., field), chrysomela.

The linkage group carrying the strongest effect QTLs, LGXIX is specific to rust resistance.

Co-localisation of QTLs for field resistance for in 4 pedigrees (8 maps).

Table 3.7: Summary of QTL results obtained in each of the 4 pedigrees studied for field rust resistance. [a] Top number is the number of QTL detected and bottom numbers are the range of PEV; [b] Percentage of explained variance.

	Pop 1		Pop2		Pop3a		Pop3b		
	TxD ♀	TxD ♂	D ♀	T ♂	D ♀	N ♂	D ♀	T ♂	Total (PEV[b] range)
Field resistance	5[a] 3.9-18.4	4 8.9-13.7	4 3.3-76.4	3 1.4-36.5	4 2.2-74.7	1 8.6	4 1.9-76.1	2 7.3-14.4	27 1.9-76.4%

Results of the analysis of co-localisation showed that:

- Six LGs are carrying QTLs that are co-localising on 2 maps (LGIII, 08, 09, 11, 12, 15);
- One LG is carrying QTLs co- localizing on 3 maps (LGVI);
- One LG is carrying QTLs - localizing on 6 maps (LGXIX).

This last linkage group is shown on Figure 3.4. Alignment with POP3 maps not possible due to the lack of microsatellites markers on the framework map used for QTL analysis. Nevertheless, this linkage group is a good candidate to further studies using genetical genomics strategy, and for two reason additional to the result presented here: LGXIX has been identify to carry qualitative resistance genes in several pedigrees [16, 25]; and genome annotation shows that this linkage group carry super –clusters of different families of Resistance Genes Analogs (RGA).

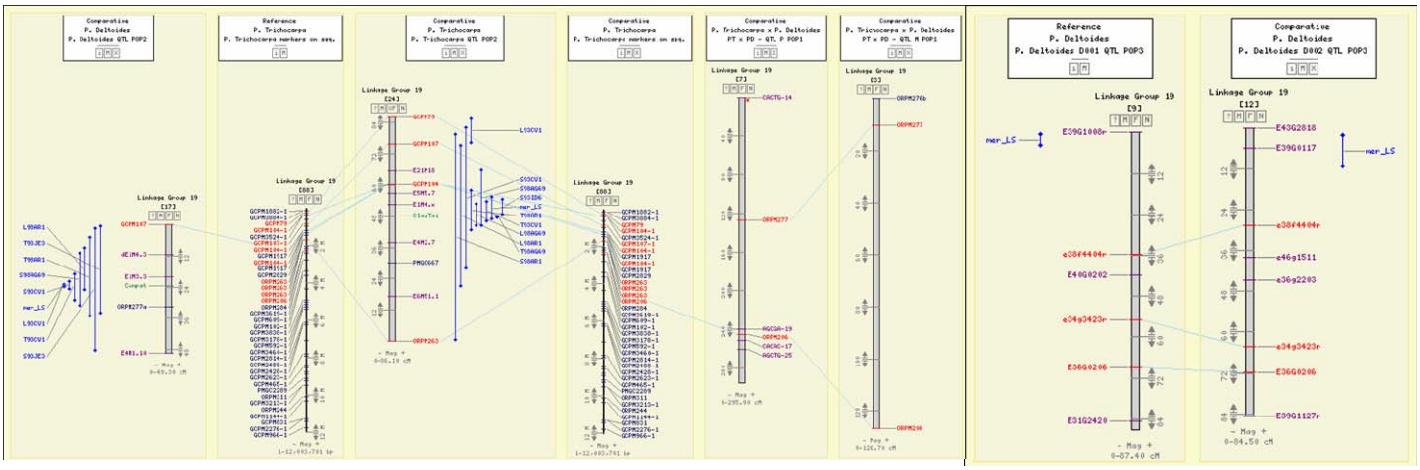


Figure 3.4 Most significant co-localisation of QTLs for rust resistance evaluated in field conditions (LGXIX), for maps of POP1, POP2 and POP3 pedigrees, and projection to genome sequence. Vertical black (genome sequence) and grey (genetic map) bars are linkage groups with markers positions; Vertical blue lines are confidence intervals of QTLs. For trait names in blue refer to previous chapters and materiel and methods. Correspondences between genetic maps and genome sequence are drawn as light blue lines.

3.4 Conclusion

This chapter presented several important and significant results:

- This is the first time that an extensive comparative QTL study across several pedigrees and environments has been carried in *Populus*, and even in trees.
- We have detected significant QTLs controlling yield and disease resistance.
- We have also detected QTLs specific to some environment and QTLs controlling this response to a specific environment.

Due to the actual moderate level of alignment of genetic maps to the genome and large confidence intervals of some QTLs detected, we could identify only large intervals often containing several hundreds of candidate genes that could underlie the QTL. To improve alignment and reduce intervals, the future strategy will consist in focusing on detected QTLs, especially those showing strong co-localisation between pedigrees. We will add more markers in these regions and, in some cases, it will be possible to increase the number of progenies analysed to reduce confidence intervals of QTLs. Finally, identification of candidate genes through an expressional approach on extreme genotypes

of the pedigree should provide an additional filter to the identification of genes underlying traits of interest.

At a longer term, this identification of genes will benefit novel strategies for selection through Gene Assisted Genotype Construction.

3.5 References

1. Bergez, JE, Auclair D, Bouvarel L. (1989) First year growth of hybrid poplar shoots from cutting or coppice origin. *Forest Science,* 35, 1105-1113

2. Barigah TS, Saugier B, Mousseau M, Guittet J, Ceulemans R. (1994) Photosynthesis, leaf area and productivity of 5 poplar clones during their establishment year. *Annals of Forest Science*, 51, 613-625

3. Heilman PE, Ekuan G, Fogle D. (1994) Above- and below-ground biomass and fine roots of 4-year-old hybrids of *Populus trichocarpa* x *Populus deltoides* and parental species in short rotation culture. *Canadian Journal of Forest Research*, 24, 1186-1192

4. Hansen E A. (1991) Poplar woody biomass yields: a look to the future. Biomass and Bioenergy*, 1, 1-7*

5. Taylor, G. (2002) *Populus: Arabidopsis* for forestry. Do we need a model tree? *Annals Bot*., 90, 681-689

6. Scarascia-Mugnozza GE, Ceulemans R, Heilman PE, Isebrands JG, Stettler RF, Hinckley TM. (1997) Production physiology and morphology of *Populus* species and their hybrids grown under short rotation. II. Biomass components and harvest index of hybrid and parental species clones. *Canadia. Journal of Forest Research*, 27, 285-294

7. Hansen EA. (1991) Poplar woody biomass yields: a look to the future. Biomass and Bioenergy*, 1, 1-7*

8. Rae AM, Robinson KM, Street NR, Taylor G. (2004) Morphological and physiological traits influencing biomass productivity in short-rotation coppice poplar. *Canadian Journal of Forest Research*, 54, 335-345

9. Lander ES and Botstein D. (1989) Mapping mendelian factors underlying quantitative traits using rflp linkage maps. *Genetics*, 121(1), 185-199

10. Haley C.S. and Knott S.A. (1992) A simple marker regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity*, 69, 315-324

11. Bradshaw HD, Stettler RF. (1995) Molecular genetics of growth and development in Populus IV. Mapping QTLs with large effects on growth, form, and phenology traits in forest trees. *Genetics*, 139, 963–973

12. Frewen BE, Chen T, Howe GT, Davis J, Rohde A, Boerjan W, Bradshaw HD. (2000) Quantitative trait loci and candidate gene mapping of bud set and bud flush in Populus. *Genetics*, 154, 837–845

13. Han, K.-H, Bradshaw, HD and Gordon, MP. (1994) Adventitious root and shoot regeneration in vitro is under major gene control in an F2 family of hybrid poplars (Populus trichocarpa x P. deltoides). *Forest Genetics*, 1(3), 139-146

14. Wu, R, Bradshaw, HD and Stettler, RF. (1997) Molecular genetics of growth and development in Populus (Salicaceae) .V. Mapping quantitative trait loci affecting leaf variation. *American Journal of Botany*, 84(2), 143-153

15. Cervera MT, Gusmao J, Steenackers M, Peleman J, Storme V. et al. (1996) Identification of AFLP molecular markers for resistance against *Melampsora larici-populina* in *Populus*. *Theoretical and Applied Genetics*, 93, 733-737

16. Newcombe, G. and Bradshaw, HD. (1996) Quantitative trait loci conferring resistance in hybrid poplar to Septoria populicola, the cause of leaf spot. *Canadian*

Journal of Forest Research-Revue Canadienne De Recherche Forestiere, 26(11), 1943-1950

17. Jorge V, Dowkiw A, Faivre Rampant P, and Bastien C. (2005) Genetic architecture of qualitative and quantitative *Melampsora larici-populina* leaf rust resistance in hybrid poplar: genetic mapping and QTL detection. *New Phytologist*, 167, 113-127

18. Tagu, D, Bastien, C, Faivre-Rampant, P, Garbaye, J, Vion, P, Villar, M. and Martin, F. (2005) Genetic analysis of phenotypic variation for ectomycorrhiza formation in an interspecific F1 poplar full-sib family. *Mycorrhiza*, 15(2), 87-91.

19. Tuskan GA, Gunter LE, Yang ZK, Yin TM, Sewell MM. and DiFazio SP. (2004) Characterization of microsatellites revealed by genomic sequencing of Populus trichocarpa. *Can. J. For. Res*, 34, 85–93

20. Lander ES, Green P, Abrahamson J, Barlow A, Daley M, Lincoln S, and Newburg L. (1987) MAPMAKER: An Interactive Computer Package for Constructing Primary Genetic Linkage Maps of Experimental and Natural Populations. *Genomics*, 1, 174-181

21. Stam P. (1993) Construction of integrated genetic linkage maps by means of a new computer package: JOINMAP. *Plant J,* 1993, 3, 739–744

22. Tuskan GA, Gunter LE, Yang ZK, Yin TM, Sewell MM. and DiFazio SP. (2004) Characterization of microsatellites revealed by genomic sequencing of Populus trichocarpa. *Can. J. For. Res,* 34, 85–93

23. Yin, T.-M, DiFazio, S P, Gunter, L E, Riemenschneider, D. and Tuskan, G A. (2004) Large-scale heterospecific segregation distortion in *Populus* revealed by a dense

genetic map. *Theor. Appl. Genet*, 109, 451–463

24. Sewell MM, Davis MF, Tuskan GA, Wheeler NC, Elam CC, Bassoni DL. and Neale D.B. (2002) Identification of QTLs influencing wood property traits in loblolly pine (Pinus taeda L.). II. Chemical wood properties. *Theoretical and Applied Genetics*, 104, 214-222

25. Grattapaglia D, Sederoff R. (1994) Genetic-linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross - mapping strategy and RAPD markers *Genetics*, 137, 1121-1137

26. Cervera MT, Storme V, Ivens B, Gusmao J, Liu B. H. et al. (2001) Dense genetic linkage maps of three *Populus* species (*Populus deltoides, P. nigra* and *P. trichocarpa*) based on AFLP and microsatellite markers. *Genetics*, 158, 787-809

27. Ritter E, Gebhardt C, Salamini F. (1990) Estimation of recombination frequencies and construction of RFLP linkage maps in plants from crosses between heterozygous parents. *Genetics*, 125, 645-654

28. Van Ooijen JW. (2004) MapQTL® 5, software for the mapping of quantitative trait loci in experimental populations. Wageningen, the Netherlands: Kyazma B.V

29. Jansen RC, Van Ooijen JW, Stam P, Lister C, Dean C. (1995) Genotype-by-environment interaction in genetic mapping of multiple quantitative trait loci. *Theoretical and Applied Genetics*, 91, 33-37

30. Romagosa, I, Ullrich, SE, Han, F. and Hayes, PM. (1996) Use of the additive main effects and multiplicative interaction model in QTL mapping for adaptation in barley. *Theoretical and Applied Genetics*, 93(1/2), 30-37

31. Dowkiw, A, Husson, C, Frey, P, Pinon, P. and Bastien, C. (2003) Partial resistance to Melampsora larici-populina leaf rust in hybrid poplars: genetic variability in inoculated excised leaf disk bioassay and relationship with complete resistance. *Phytopathology*, 93(4), 421-427

32. Barigah TS, Saugier B, Mousseau M, Guittet J, Ceulemans R. (1994) Photosynthesis, leaf area and productivity of 5 poplar clones during their establishment year. *Annals of Forest Science*, 51, 613-625

33. Heilman PE, Ekuan G, Fogle D. (1994) Above- and below-ground biomass and fine roots of 4-year-old hybrids of *Populus trichocarpa* x *Populus deltoides* and parental species in short rotation culture. *Canadian Journal of Forest Research*, 24, 1186-1192

34. Hansen E A. (1991) Poplar woody biomass yields: a look to the future. Biomass and Bioenergy*, 1, 1-7*

35. Taylor, G. (2002) *Populus: Arabidopsis* for forestry. Do we need a model tree? *Annals Bot*., 90, 681-689

36. Scarascia-Mugnozza GE, Ceulemans R, Heilman PE, Isebrands JG, Stettler RF, Hinckley TM. (1997) Production physiology and morphology of *Populus* species and their hybrids grown under short rotation. II. Biomass components and harvest index of hybrid and parental species clones. *Canadia. Journal of Forest Research*, 27, 285-294

37. Hansen EA. (1991) Poplar woody biomass yields: a look to the future. Biomass and Bioenergy*, 1, 1-7*

38. Rae AM, Robinson KM, Street NR, Taylor G. (2004) Morphological and physiological traits influencing biomass productivity in short-rotation coppice poplar. *Canadian Journal of Forest Research*, 54, 335-345

39. Lander ES and Botstein D. (1989) Mapping mendelian factors underlying quantitative traits using rflp linkage maps. *Genetics*, 121(1), 185-199

40. Haley C.S. and Knott S.A. (1992) A simple marker regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity*, 69, 315-324

41. Bradshaw HD, Stettler RF. (1995) Molecular genetics of growth and development in Populus IV. Mapping QTLs with large effects on growth, form, and phenology traits in forest trees. *Genetics*, 139, 963–973

42. Frewen BE, Chen T, Howe GT, Davis J, Rohde A, Boerjan W, Bradshaw HD. (2000) Quantitative trait loci and candidate gene mapping of bud set and bud flush in Populus. *Genetics*, 154, 837–845

43. Han, K.-H, Bradshaw, HD and Gordon, MP. (1994) Adventitious root and shoot regeneration in vitro is under major gene control in an F2 family of hybrid poplars (Populus trichocarpa x P. deltoides). *Forest Genetics*, 1(3), 139-146

44. Wu, R, Bradshaw, HD and Stettler, RF. (1997) Molecular genetics of growth and development in Populus (Salicaceae) .V. Mapping quantitative trait loci affecting leaf variation. *American Journal of Botany*, 84(2), 143-153

45. Cervera MT, Gusmao J, Steenackers M, Peleman J, Storme V. et al. (1996) Identification of AFLP molecular markers for resistance against *Melampsora larici-populina* in *Populus*. *Theoretical and Applied Genetics*, 93, 733-737

46. Newcombe, G. and Bradshaw, HD. (1996) Quantitative trait loci conferring resistance in hybrid poplar to Septoria populicola, the cause of leaf spot. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere*, 26(11), 1943-1950

47. Jorge V, Dowkiw A, Faivre Rampant P, and Bastien C. (2005) Genetic architecture of qualitative and quantitative *Melampsora larici-populina* leaf rust resistance in hybrid poplar: genetic mapping and QTL detection. *New Phytologist*, 167,

113-127

48. Tagu, D, Bastien, C, Faivre-Rampant, P, Garbaye, J, Vion, P, Villar, M. and Martin, F. (2005) Genetic analysis of phenotypic variation for ectomycorrhiza formation in an interspecific F1 poplar full-sib family. *Mycorrhiza*, 15(2), 87-91.

49. Tuskan GA, Gunter LE, Yang ZK, Yin TM, Sewell MM. and DiFazio SP. (2004) Characterization of microsatellites revealed by genomic sequencing of Populus trichocarpa. *Can. J. For. Res*, 34, 85–93

50. Lander ES, Green P, Abrahamson J, Barlow A, Daley M, Lincoln S, and Newburg L. (1987) MAPMAKER: An Interactive Computer Package for Constructing Primary Genetic Linkage Maps of Experimental and Natural Populations. *Genomics*, 1, 174-181

51. Stam P. (1993) Construction of integrated genetic linkage maps by means of a new computer package: JOINMAP. *Plant J,* 1993, 3, 739–744

52. Tuskan GA, Gunter LE, Yang ZK, Yin TM, Sewell MM. and DiFazio SP. (2004) Characterization of microsatellites revealed by genomic sequencing of Populus trichocarpa. *Can. J. For. Res,* 34, 85–93

53. Yin, T.-M, DiFazio, S P, Gunter, L E, Riemenschneider, D. and Tuskan, G A. (2004) Large-scale heterospecific segregation distortion in *Populus* revealed by a dense genetic map. *Theor. Appl. Genet*, 109, 451–463

54. Sewell MM, Davis MF, Tuskan GA, Wheeler NC, Elam CC, Bassoni DL. and Neale D.B. (2002) Identification of QTLs influencing wood property traits in loblolly pine (Pinus taeda L.). II. Chemical wood properties. *Theoretical and Applied Genetics*, 104, 214-222

55. Grattapaglia D, Sederoff R. (1994) Genetic-linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross - mapping strategy and RAPD markers *Genetics*, 137, 1121-1137

56. Cervera MT, Storme V, Ivens B, Gusmao J, Liu B. H. et al. (2001) Dense genetic linkage maps of three *Populus* species (*Populus deltoides, P. nigra* and *P. trichocarpa*) based on AFLP and microsatellite markers. *Genetics*, 158, 787-809

57. Ritter E, Gebhardt C, Salamini F. (1990) Estimation of recombination frequencies and construction of RFLP linkage maps in plants from crosses between heterozygous parents. *Genetics*, 125, 645-654

58. Van Ooijen JW. (2004) MapQTL® 5, software for the mapping of quantitative trait loci in experimental populations. Wageningen, the Netherlands: Kyazma B.V

59. Jansen RC, Van Ooijen JW, Stam P, Lister C, Dean C. (1995) Genotype-by-environment interaction in genetic mapping of multiple quantitative trait loci. *Theoretical and Applied Genetics*, 91, 33-37

60. Romagosa, I, Ullrich, SE, Han, F. and Hayes, PM. (1996) Use of the additive main effects and multiplicative interaction model in QTL mapping for adaptation in barley. *Theoretical and Applied Genetics*, 93(1/2), 30-37

61. Dowkiw, A, Husson, C, Frey, P, Pinon, P. and Bastien, C. (2003) Partial resistance to Melampsora larici-populina leaf rust in hybrid poplars: genetic variability in inoculated excised leaf disk bioassay and relationship with complete resistance. *Phytopathology*, 93(4), 421-427

62.

4. Linkage disequilibrium and association mapping for phenological traits in *Populus*

M Morgante, W Boerjan, G Zaini, G Taylor,C Bastien, R Ceulemans, N Marron, S Dillen, A Rohde, T Ruttink, V Storme

4.1 Introduction

Genetic variation exists within every species and forms the basis for selection and evolution. Levels and patterns of genetic diversity vary greatly within and among populations. Genetic markers, like isozymes, random amplified polymorphic DNA (RAPD), microsatellites (SSR), amplified fragment length polymorphisms (AFLP) have been widely used to estimate that variation and to describe related phenomena, such as gene flow, effective population size and historical processes that have moulded the genetic structure of the species [1]. Compared to the above mentioned markers, direct DNA sequence analysis provides the highest level of genetic resolution and additionally provides information that can be used in phylogenetic analysis. Therefore, thanks to improvements in sequencing technology and the development of high-throughput genotyping methods, the detection of nucleotide polymorphisms (SNP: single nucleotide polymorphism, INDEL: insertion/deletion) is becoming the primary mean to analyse naturally occuring allelic variation and detect associations between the allelic forms of a gene and phenotypes [2]. Association studies, that make use of historical recombinants rather than ad hoc created ones, allow a shorter time to trait mapping, because no mapping population need be created, and higher resolution than traditional map-based strategies because of a larger number of meiotic recombination events. Thus, association strategies would be very effective in outcrossing long-lived plants, where the long generation times and the difficulties in obtaining segregating populations make other mapping strategies difficult [1].

There are two different approaches to perform association studies: a direct one involves the evaluation of candidate gene diversity across individuals from natural populations and its association with phenotypic variations [3]; a second indirect one requires a whole-

genome scan through a dense map of SNPs. The first one is not so much a mapping strategy, because it basically allows one to discard genes that do not show a significant association and to further test the remaining ones to define a cause effect relationship. The second one is a mapping strategy since it doesn't necessarily lead to the identification of the causative gene but rather of more or less closely linked polymorphisms (depending on the level of linkage disequilibrium in the region and in the population). Essential elements to define the association strategy to undertake are the detection of polymorphic single nucleotide markers and the estimation of the linkage disequilibrium (LD) extent and structure in the genome under study [4].

To date in plants, genetic diversity at the sequence level and LD structure have been studied in only a few taxa, with the maximum efforts in maize and Arabidopsis. Maize is considered highly polymorphic, with a high frequency of nucleotide changes [5](1 polymorphism per 31 bp in non-coding regions and 1 per 124 bp in coding regions); and a wide range of nucleotide diversity (p: 0.00047 for the promoter of tb1, [6]; p: 0.0173 for all sites and 0.037 for synonymous sites at glb1 locus, [7]); different surveys suggest that LD structure decays quite rapidly (within few hundred bases) in different maize populations, while it is possible to define populations with strong LD [8]. *A. thaliana*, mainly a selfing species, usually harbors extensive LD (it decays within 1 cM or 250 kb, [9]) and has generally lower levels of diversity than maize: at the FRIGIDA gene p was found to be equal to 0.0021 for all sites and 0.0027 for synonymous sites [10], at the PgiC locus p resulted to be equal to 0.0038 for all sites and 0.0051 for synonymous sites [11]. When considering woody perennials, data on nucleotide polymorphisms are available for *Pinus sylvestris* (p= 0.0049 for synonymous sites of PAL gene; [12]; (p= 0.0013-0.0024 for synonymous sites of two phytochrome genes; [13]), *Picea abies* (p= 0.0063 for synonymous sites; degli Ivanissevich & Morgante, (unpublished data) and *Betula pendula* (p= 0.0043 for synonymous sites of MADS2 locus; [14]).

In *Populus nigra*, a common pioneer species of the riparian forest in Europe and western Asia, genetic diversity has been assessed through allozyme variation [15]), AFLP

markers [16, 17] and nuclear microsatellites [18], where a high level of diversity and significant differentiation among populations were found. A more recent study employed isozymes, microsatellites and AFLPs to look at genetic variation in gene bank collections from nine different European countries and detected high variation within and moderate differentiation among geographical regions [19].

Recently, estimates of nucleotide polymorphism frequency in *Populus* and other forest trees have been completed and phytochrome genes identified as important for phonological adaptation in aspen [20, 21].

4.2 Materials and Methods

Assessment of the bud set phenotype in the *P. nigra* population

Bud set was scored according to a new bud-set phenology scorecard that was developed as described in Plate 4.1 The Gent team trained other partners in the use of this scorecard and they have employed it for bud set description in the POP families. Phenotypic bud set data were collected during a period of 84 days, from August 22 till November 14, 2005. Trees were scored twice a week during this period (binding on average four people/day). This intense scoring allowed us to capture the full genetic variation present in the *Populus nigra* collection whose origins span about 9 degree's latitude. The obtained data have a high resolution and allow for the first time the description of the dynamics of autumnal growth arrest and dormancy induction. Most previous phenotypic measurements have considered the day of completed bud set; such data have been used in all previous QTL detection. However, the date of onset of the process as well as the pace of it might identify other and more trait-linked QTLs.

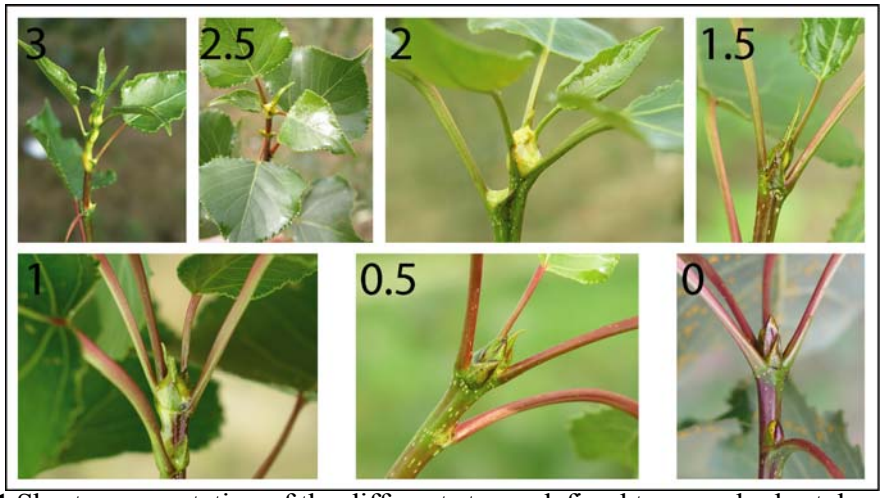


Plate 4.1 Short representation of the different stages defined to score bud set, here with photographs from *P. nigra*.

The buds of more than 2500 *P. nigra* trees were evaluated. Tree height was assessed on August 22, and again on December 19 in order to compare to a previously employed method for inferring date of bud set through the amount of late-seasonal height increment [22]. Moreover, trees with a potential bias on bud set through their architecture were categorized to evaluate this confounding factor.

Identification of a set of candidate genes for bud set, including support from QTL detection and transcriptome analysis

Literature search for candidate genes for bud-set

We generated a list of 37 plant genes reported to be involved in light signal reception and transduction as well as in light- and temperature-dependent control of flowering time, this list including phytochrome and cryptochrome genes, phytochrome signal transduction genes, circadian clock genes, vernalization genes and finally key-genes responsible for the integration of different light, temperature and growth signals upstream of floral homeotic genes. Most of those accessions derive from *Arabidopsis thaliana* since a

consistent number of mutants impaired in circadian clock function and light-mediate reproductive rhythms were produced and well documented for this model organism. Nevertheless crop plants sequences whose implication in flowering time determination were either hypothesized or supported with experimental evidence (e.g. *Heading date* from rice , *dwarf8* and *indeterminate1* from maize, ZGT from tobacco) were also included when available in GenBank databases. Following the acquisition of coding sequences from GenBank accessions, we performed TBLASTN analyses versus nr and EST databases in order to find homologous sequences in the *Populus* genus. The same type of analysis was performed by comparing those sequences with a genomic database of more than 1.7 Gbases available at DOE Joint Genome Institute website (http://genome.jgi-psf.org/poplar0/poplar0.home.html). Best score hits were recorded by taking in consideration BLAST score value, E-value and total matching peptide length. See attached report and tables for a full description of the information available.

Gene	Gene full name	Organism
PHYA	Phytochrome A	*Arabidopsis thaliana*
PHYB	Phytochrome B	*Arabidopsis thaliana*
PHYC	Phytochrome C	*Arabidopsis thaliana*
PHYD	Phytochrome D	*Arabidopsis thaliana*
PHYE	Phytochrome E	*Arabidopsis thaliana*
PAP1	Phytochrome-associated protein 1	*Arabidopsis thaliana*
PAP2	Phytochrome-associated protein 2	*Arabidopsis thaliana*
PAP3	Phytochrome-associated protein 3	*Arabidopsis thaliana*
PIF3	Phytochrome-interacting factor 3	*Arabidopsis thaliana*
PIF4	Phytochrome-interacting factor 4	*Arabidopsis thaliana*
HFR1 / REP1	Long hypocotyl in far-red 1 / Reduced phytochrome signaling 1	*Arabidopsis thaliana*
PAT1	Phytochrome A signal transduction 1 protein	*Arabidopsis thaliana*
PAT3	Phytochrome A signal transduction 3 protein	*Arabidopsis thaliana*
SPA1	Phytochrome A supressor spa1	*Arabidopsis thaliana*
PSK1	Phytochrome kinase substrate 1	*Arabidopsis thaliana*
PIM1	Phytochrome interacting molecule 1	*Arabidopsis thaliana*
ADO1 / ZTL / LKP1	Adagio 1 / Zeitlupe / LOV kelch protein 1	*Arabidopsis thaliana*
ADO2 / LKP2	Adagio 2 / LOV kelch protein 2	*Arabidopsis thaliana*
ADO3 / FKF1	Adagio 3	*Arabidopsis thaliana*
CCA1	Circadian clock associated 1	*Arabidopsis thaliana*
LHY	Late elongated hypocotyl	*Arabidopsis thaliana*
TOC1	Timing of CAB expression 1	*Arabidopsis thaliana*
ELF3	Early flowering 3	*Arabidopsis thaliana*
GI	Gigantea	*Arabidopsis thaliana*
CK2	Regulatory subunit of protein kinase CK2	*Arabidopsis thaliana*
Co	Constans	*Arabidopsis thaliana*
FT	Flowering locus T	*Arabidopsis thaliana*
SOC1	Suppressor of Constans overexpression 1	*Arabidopsis thaliana*
TFL1	Terminal flower 1	*Arabidopsis thaliana*
TFL2	Terminal flower 2	*Arabidopsis thaliana*
VRN2	Vernalization 2	*Arabidopsis thaliana*
FRI	Frigida	*Arabidopsis thaliana*
FLC	Flowering locus C	*Arabidopsis thaliana*
ZGT	Circadian clock coupling factor ZGT	*Nicotiana tabacum*
HD1	Heading date 1	*Oryza sativa*
Dwarf8	Dwarf8	*Zea mays*
ID1	Indeterminate 1	*Zea mays*

Table 4.1 Candidate genes for bud set

Among the most significative hits in **GenBank nr database**, the following sequences were found:

Query	GenBank Accession	GenBank accession gi
PHYA	Populus tremula x Populus tremuloides mRNA for phytochrome A	2664188
PHYB	Populus balsamifera subsp. trichocarpa phytochrome B1 (phyB1) gene,complete cds	10954090
PHYB	Populus balsamifera subsp. trichocarpa phytochrome B2 (phyB2) gene,complete cds	10954092
PAP1[1]	Populus tremula x Populus tremuloides mRNA for aux/IAA protein (IAA7 gene)	20269060
PAP2[1]	Populus tremula x Populus tremuloides mRNA for aux/IAA protein (IAA2 gene)	20269050
SOC1[2]	Populus tremuloides SEP3-related MADS-box protein (M6) mRNA, complete cds	30314023
FLC	Populus tremuloides transcription factor MAGL4 mRNA, complete cds	28372801

The transcriptome in apical buds during dormancy as visualized by cDNA-AFLP

Gene expression in apical buds during this developmental process was monitored using cDNA-AFLP transcript profiling [23]. This quantitative profiling approach allows the simultaneous monitoring of principally all genes, given the limitation that a particular cDNA must have restriction sites for the two restriction enzymes chosen (BstYI, MseI) and that the subsequently amplified fragment has a length that can be visualized in polyacryl amide gels. An *in silico* analysis of *Arabidopsis* genes demonstrated that the method can visualize approximately 60% of all genes using one particular restriction enzyme combination. A similar degree of transcriptome resolution is assumed for poplar.

120 of the 128 possible primer combinations gave informative results and were scored for interesting fragments. By eye, three to four different big clusters of gene expression were noticed in the expression profiles throughout time. During the first period of dormancy induction, growth is still occurring, though at an ever-decreasing rate. The expression profiles of the corresponding samples differ clearly from all later samples. Many genes become expressed after growth has arrested, and might at least in part correlate with genes expressed in bud scales. The expression of quite a number of these genes is interrupted in the middle, a period most likely corresponding to dormancy. However, there are other alterations in gene expression during dormancy itself. These trends in gene

expression have been observed with all primer combinations used to subsequently amplify all subpools of the cDNA mix.

These gross changes in gene expression appeared not to correlate with the time points where light and temperature conditions were adjusted. The phenological observations (*e.g.*, bud scales visible, or bud set) do neither coincide with one of the big groups of gene expression. The only plausible conclusion is that these changes in gene expression are inherent to the process of bud development. Taking into account that dormancy is still very poorly described at the molecular level we decided to analyse the following tags in more detail: i) tags with an expression during a longer period of time corresponding either to physiological stages or to one of the big expression clusters, ii) tags expressed during one or two timepoints preceding the expression of the other groups. Differentially expressed tags were re-amplified and directly sequenced (currently 470 tags sequenced).

Identity of genes that are differentially expressed

In the second step, the sequences were assigned a putative function via semi-automated database searching. All fragments were compared with the protein database (BlastP), with the available poplar ESTs (BlastN) and the poplar genome (BlastN). It is inherent to the cDNA-AFLP method used that mainly 3'sequences are generated. If these fragments have no or only little coding sequence, often nothing is found in the databases. For these fragments, we are currently developing a method to create a longer sequence from a match with the genome sequence, so to repeat the search against the protein database with a longer sequence. This match with the poplar genome sequence will at the same time link the candidate genes with the QTLs controlling the natural variation in dormancy-related traits (bud set, bud flush).

Candidate genes for bud-set from cDNA-AFLP and transcriptomic approaches

The dates of bud flush and bud set delimit the productive growth period of trees. Growth alternates with seasonal dormancy that lasts in the temperate climate about 6 months or longer. The whole process of dormancy induction (bud set), dormancy and dormancy release (bud flush) can be mimicked in controlled growth chamber experiments. Key in

these experiments are the changes in light and temperature: a short day of 8h together with declining temperature is applied to induce dormancy; a treatment at 4°C fulfills the chilling requirement, and plants are finally transferred to long days to resume growth.

Over the experimental period of 115 days, samples of apical buds of *Populus tremula* x *P. alba* were collected at 33 points in time. Apex/bud-derived RNA was reverse-transcribed into cDNA and changes in gene expression were monitored using cDNA-AFLP [23]. Taking into account that bud development and dormancy is still very poorly described at the molecular level, we decided to analyse the following tags in more detail: i) tags with an expression during a longer period of time corresponding either to physiological stages or to one of the big expression clusters, ii) tags expressed during one or two timepoints preceding the expression of the other groups. A total of 483 differentially expressed tags were re-amplified and directly sequenced. These tags were clustered into groups according to their temporal expression pattern.

One third of these sequences could be matched to the current version of the poplar genome sequence (http://genome.jgi-psf.org/Poptr1/Poptr1.home.html). A surprising large fraction is possibly of bacterial origin, given that a number of tags have strong hits to bacterial sequences. These fragments are not expressed during bud development, but only during the later stages of dormancy and might thus be derived from a specific infection or an endophytic bacterium that persists specifically in buds. Yet, another 184 transcript-derived fragments have no hits at all to the database. These fragments have generally no coding potential and are probably derived from 3' sequences, for which the cDNA-AFLP technique enriches. Our gene identification is undertaken in *Populus tremula* x *P. alba*, a species from a different section in the genus *Populus* as compared to the sequenced *Populus trichocarpa* 'Nisqually'. Thus, it is possible that the corresponding genes are not found due to the diversity of 3' sequences between these two species.

Nevertheless, 160 poplar genes could be mapped onto the process and allowed to reconstruct a number of signal transduction and biochemical pathways to be described.

Changes in the carbohydrate metabolism occur early during the process of bud development. Genes encoding components of ß-oxidation of fatty acids, osmoprotectants or late-embryogenesis abundant proteins are abundantly present during dormancy and comply with earlier physiological studies. Also components of signal transduction pathways were identified, among others for ethylene and auxin, but not for abscisic acid. Integrated with the phenology and morphology of the process, a description of molecular events is emerging.

Assessment of nucleotide diversity within the *P. nigra* population

Twelve naturally occurring *Populus nigra* plants were included in the study (Table 4.2). Ten plants were obtained from I.N.R.A. – Unité Amélioration Génétique et Physiologie Forestières (Orleans, France) and the other two plants from DiSAFRi -University of Tuscia (Viterbo, Italy). Each of the individuals used originated from a different and geographically distant population from the collections described in Chapter 1. Geographic locations of the 12 plants are provided in Table 4.2. Leaf tissue from field-grown clones was collected for DNA extraction.

No.	Name	Location
1	BDG (Blanc de Garonne)	Garonne river (F), exact location unknown
2	71077-308	Ain river, Dept. Ain (F), Lat. 45.55° N, Long. 5.14° E
3	72131-03	Allier river, Dept. Puy de Dome (F), Lat. 45.37° N, Long. 3.13° E
4	72145-7	Gard river, Dept. Gard (F), Lat. 43.59° N, Long. 4.13° E
5	71030-501	Isère river, Dept. Savoie (F), Lat. 45.35° N, Long. 6.27° E
6	SRZ (Sarrazin)	Lot river, Dept. Lot (F), Lat. 44.28° N, Long. 1.26° E
7	92510-1	Loire river, Dept. Nièvre (F), Lat. 47.28° N, Long. 2.54° E
8	73193-84	Gave de Gavarnie river, Dept. Hte Pyrénees, Lat. 42.57° N, Long. 0.04° E
9	52-27	Dept. Ain (F), exact location not available
10	71073-305	Rhone river, Dept. Ain (F), Lat. 45.47° N, Long. 5.27° E
11	Poli	Bosco di Policoro, Matera (I), Lat. 40°N, Long. 16°E
12	58-861	Venaus, Torino (I), Lat. 45°N, Long. 7°E

Table 4.2 Location of origin of the individuals used to assess nucleotide diversity

DNA extraction

Leaf material (fresh or frozen at –80 °C) was ground into a fine powder using mortar and pestle in the presence of liquid nitrogen. The genomic DNA was then extracted using DNeasy Plant kit (Qiagen, Inc. Valencia, CA), according to the standard protocol provided by the manufacturer.

Gene sequences and primer design

A total of 31 genomic regions were amplified on the 12 different genotypes. Sequence specific primer pairs for the PCR were designed using the PRIMER3 software (http://www.genome.wi.mit.edu). The expected size of the amplification fragments was chosen to range from 400 to 800 bp. Three adjacent regions were selected from phytochrome A (PHYA, gi: 2664188) and other two from auxin/IAA2 (IAA2, gi: 20269050) cDNAs of *Populus tremula* x *Populus tremuloides*, where the presence of introns was hypotetically predicted on the basis of the similarity to the related genes of *A. thaliana*. One region was selected from phytochrome B1 (PHYB1, gi: 10954090) and two non-overlapping regions from phytochrome B2 (PHYB2, gi: 10954092) complete gene sequences from *Populus balsamifera*. In order to enable the amplification of those regions in P. nigra, primer design was targeted to exon sequences. Also cryptochrome 1 (CRY1, gi:442528) and cryptochrome 2 (CRY2, gi:1514442) from *A. thaliana* were considered. Since no cryptochrome-like sequences belonging to the genus Populus were present in GeneBank, the sequences of *A. thaliana* were blasted against the genome database of P. trichocarpa (available at http://genome.jgi-psf.org/poplar0/poplar0.home.html) consisting of single unassembled sequence reads. Sequences showing similarity (E < 1050) to CRY1 and CRY2 were assembled using the CAP3 software [24], in order to obtain genomic contigs containing putative cryptochrome sequences. Specific primers were designed on the basis of the contig consensus. Similarly, a third region for IAA2 was selected on a contig obtained from blast analysis of IAA2 cDNA (gi: 20269050) against the *Populus* genomic sequences. The other 20 regions, representing random single-copy sequences in the *P. nigra* genome, were identified by sequencing the ends of the inserts of a genomic BAC library of *P.*

nigra x *P. deltoides* (provided by Dept. of Plant Systems Biology, VIB, University of Ghent, Belgium). BAC DNA was purified from 576 clones using Whatman 96-well Unifilters and Uniplates (Whatman, Clifton, NJ). BAC insert extremities (BAC ends) were directly sequenced using M13 universal forward and reverse primers and the ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction kit (Applied Biosystems, Foster City, CA). The sequences, analysed on an ABI3730 sequencer (Applied Biosystems, Foster City, CA), were blasted against GenBank database and the genome database of *P. trichocarpa* in order to select those pairs of BAC ends (M13 forward sequence and M13 reverse sequence of the same BAC insert) that putatively represent single-copy sequences. The criteria for the selection were the following: no blastN and blastX hits against the GenBank nr database with E< e-20 (this allowed to discard sequences similar to mithocondrial DNA, chloroplast DNA, and known transposable elements); less than 16 blastN and blastX hits against the P. trichocarpa genome database (representing approximately 8 genome equivalents) with E< e-20 (this allowed to discard sequences containing putative repetitive elements within *Populus nigra*).

PCR amplification

DNA amplifications were performed in a 25 µl volume. The reactions contained 25 ng of genomic DNA, 0.3 µM of each primer, 250 µM of each dNTP, 1.5 mM MgCl2, 5% DMSO, 1.25 units AmpliTaq Gold (Applied Biosystems, Foster City, CA) and 1X AmpliTaq Gold buffer. The reactions were performed in the GeneAmp 9700 PCR system (Applied Biosystems, Foster City, CA), under the following conditions: 94 °C for 10 min., 35 cycles of 15 sec. at 94 °C, 30 sec. at 50 °C and 1 min. at 72 °C, followed by a final extension of 7 min. at 72 °C. PCR products were analysed on agarose gel and purified using the PCR96 Cleanup plates (Millipore, Bedford, MA) with a vacuum manufold, according to the manufacturer's instructions. Prior to DNA sequencing, the amount of the purified products was estimated on agarose gel.

DNA sequencing

Purified PCR products were sequenced directly on both strands using the locus specific primers and the ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction kit v3.1, then separated on an ABI3730 sequencer. The output sequences of each region for all the genotypes were trimmed using Lucy software [25] and aligned and visualised using Phred-Phrap-PolyPhred-Consed programs [26, 27, 28].

Data analysis

Base changes (SNP) and insertion/deletion polymorphisms (indel) were identified directly by visual inspection of sequence alignments after the PolyPhred SNP detection and catalogued in an Excel spreadsheet. The number of transitions and transversions were computed. Frequencies of polymorphic sites (per 100 bp) were calculated by dividing the total number of polymorphic sites by the length of the DNA sequence examined. Haplotypes were inferred by using Haplotyper program [29]. Genetic parameters (genetic diversity p, Tajima D) and synonimous/non synonimous changes were estimated using DNAsp program [30]. Linkage disequilibrium (LD) measures D' and R2 were calculated with DNAsp for SNPs within each locus/sequence using the haplotype information obtained from Haplotyper, and with LD Analyzer [31]; available at http://www.chgb.org.cn/lda/lda.htm) for SNPs between different loci/sequences.

Association between budset phenotype and candidates genes

The genotyping analysis was performed using two techniques both based on the single base extention approach (SBE). A SBE genotyping analysis was developed through a Template-directed Dye-terminator Incorporation with Fluorescence Polarization detection (TDI-FP technique) using the AcycloPrime II™ Detection Kit and a FP reader (Victor^2) from PerkinElmer. The technique proved to be cheaper and faster than other available techniques, but not as sensitive as expected providing a large number of negative data (Fig. 4.1A). To confirm data when necessary, we also adopted the fluorescence dye-labelled SBE reaction performed with ABI PRISM® SNaPshot™ Kit and detected with the Sequencer 3730 (from Applera). This second technique resulted more sensitive, even

if more expensive and time consuming (Fig. 4.1B). The genotyping analysis was performed on 398 DNA samples (Table 4.3), which belong to the wider LD population established in Belgium. The remaining 126 out of 524 clones had to be excluded for the moment from the association analysis as: i. hybrid clones (as in the French populations); ii. mixed genotypes among blocks (as for Spanish clones); iii. clones with less than 3 phenotypic measures for bud set trait (< 3 ramets/blocks); iv. clones with leaf material missing. Concerning the 86 Spanish clones which belong to mixed genotypes, it is planned to genotype the samples available and the trees in the Belgian field in order to understand which genotype is which and rescue them for the association study.

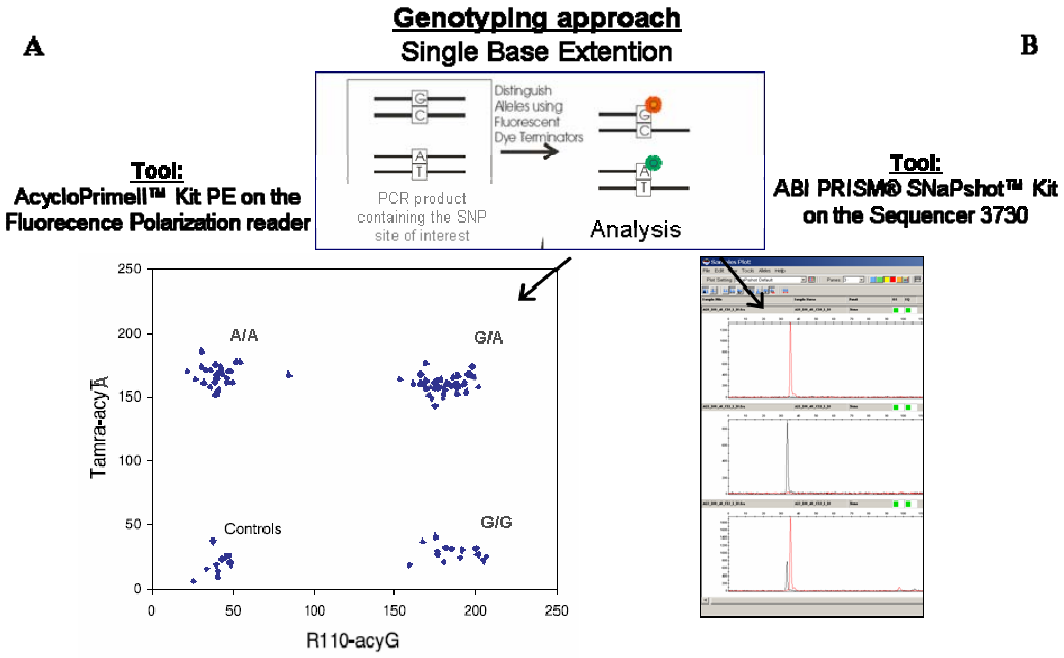


Figure 4.1 Output of the genotyping analysis performed by P9 using: **A** the AcycloPrime II™ Detection Kit and a FP reader (Victor^2) from PerkinElmer; **B** ABI PRISM® SNaPshot™ Kit and the Sequencer 3730 from Applera.

Country	Initial clones	DNA samples	Clones to genotype	Latitude (° North)
The Netherlands	50	49	49	52.16 – 52.37
Germany	56	56	50	49.49
France	192	191	170	44.00 – 47.28
Italy	107	107	102	45.12 – 45.16
Spain	119	75	27	41.35 – 41.55
Total	**524**	**478**	**398**	41.35 – 52.37
Bridges	20	8	8	

Table 4.3 Material available for the association analysis

For the association analysis 9 candidate genes (Table 4.4) out of the 37 ones analysed during the preliminary study were identified, focusing on those phytochromes, cryptochromes, signal integrators, vernalization factors and circadian clock factors suggested as more interesting by the recent literature and by current research. A total of 27 informative SNPs (TagSNP in the Table 4.4) were successfully genotyped; the individuals used for the tagSNP discovery were also included in each genotyping experiment as positive controls. The genotypic data obtained were preliminary used to compute the F-statistics (estimated as average of Fst of each SNP considered as an independent locus) using the software Arlequin 2.0 [32]. Low-to-moderate levels of genetic differentiation (Fst from 0.05 to 0.1; total Fst= 0.06) were observed in the candidate genes analysed (Fig. 4.1), even if the distribution of Fst values for each polymorphic locus may suggest the presence of few outlying SNPs (Fst > 0.15).

The SNP genotypic data of the candidate genes were then used to perform association analysis, using the software TASSEL (Trait Analysis by Association, Evolution and Linkage; http://www.maizegenetics.net/tassel). Phenotypic data for the association were as described in 4.2.1. The association was computed by a least square analysis according to the General Linear Model tool provided by the software, where the SNP genotypic data were considered as the only effect on the model and the significance of marker effect was expressed using 1000 permutations. At the moment no structure and kinship data were included in the model as not still available.

Candidate Gene	Resequenced product	TagSNP genotyped
PHYA gi:2664188 AJ001318 phytochrome A	*PHYA-100* 572 bp	4
	PHYA-102F/R 501 bp	
PHYB1 gi:10954090 AF309806 phytochrome B1	*PHYB1-104* 589 bp	1
PHYB2 gi:10954092 AF309807 phytochrome B2	*PHYB2-108* 545 bp	7
	PHYB2-110 611 bp	
CRY gi:442528 cryptochrome	*CRY1-128* 427 bp	1
CRY1 *(LGII)*	*CRY1185* 630 bp	
	CRY1187 651 bp	
CRY2 *(sc273)*	*CRY2181* 580 bp	2
	CRY2183 621 bp	
CRY4 *(LGV)*	*CRYV197* 551 bp	1
CoL NM_121589 gi:18417804 Constans-like 1	*Co135* 447 bp	4
	Co137 535 bp	
	Co470 381 bp	
FRI gi:10801175 AF228500 Frigida	*FRI139* 636 bp	3
	FRI141 372 bp	
PAT1 gi: 8132288	*PAT1P199* 671 bp	4
Phy A signal transduction		

Table 4.4 List of candidate genes and SNP loci genotyped in the LD population

4.3 Results

Phenotypic variation in bud set within *P. nigra*

As shown in Figure 4.2, *P. nigra* accessions are resolved in their latitudinal gradient; northern clones will start the process of bud set earlier than southern clones, when grown together at the field site in Belgium (about 33 days of difference for the onset of the

process). This observation fits the expectation that the critical day length, varying with the latitude of origin, governs a major part of the phenotype. After the onset of the process, however, different stages or subprocesses defined in our scoring system last for varying periods of time in the accessions (compare the slope of the curves in Figure 4.2). Interestingly, genotypes at the phenotypic ends change with subprocess, meaning that different genetic components control each of the subprocesses.

These phenotypic data, together with those data recorded with the same scoring system in the POP pedigrees, will enable us to define subprocesses, their phenotypic range and their respective genetic denominators.

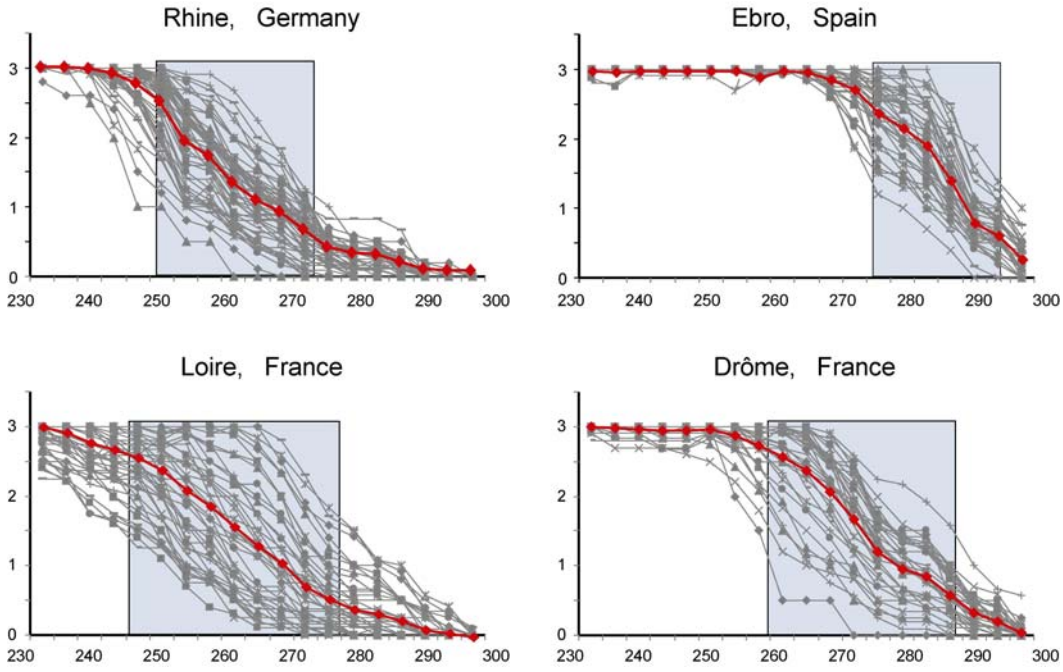


Figure 4.2 Examples of the evolution of bud set in accession from different geographical regions. Stages of bud development (3 to 0) are expressed in function of Julian days; the red line depicts the mean of genotypes included.

Candidate genes for bud-set from transcriptome analysis

A total of 1484 expressional candidate genes have been identified for bud set by P2 using two transcriptomic approaches, namely cDNA-AFLP and microarray, the latter in collaboration with Rishi Bhalerao at UPSC. These experiments were run in controlled conditions and have a high temporal resolution to follow the process of bud development and bud set.

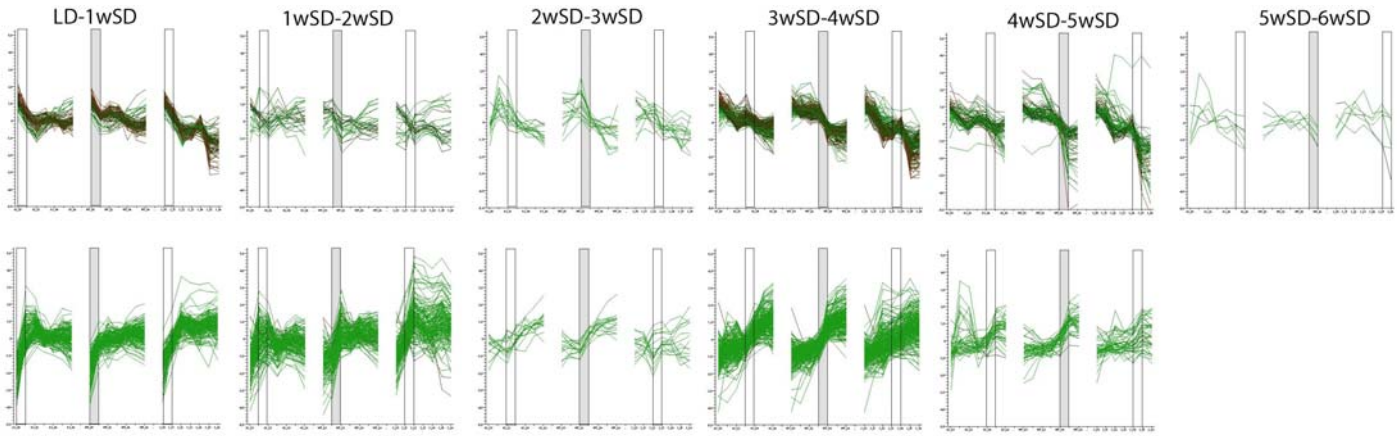


Figure 4.3 Temporal categorization of expression clusters. Expression profiles are categorized for the time, when the highest change (down-regulation, upper panel; up-regulation, lower panel) in expression is achieved. Each panel shows in the middle, the expression profile of wild type.

Each panel shows in the middle, the expression profile of wild type. The genes from both the microarray and the cDNA-AFLP experiments have been clustered into temporal groups, as exemplified in Figure 4.3. These temporal clusters contain also genes of similar functions or highlight pathways that are active during a specific stage. Transcription factors with typical expression profiles are currently being analyzed further.A total of 1188 genes (microarray: 1071, cDNA-AFLP: 132, common: 15) show an interesting gene expression throughout the process and were placed on the genome sequence. Of these, 955 genes fall onto linkage groups of the assembled poplar genome sequence, another 233 genes are

contained on unassembled scaffolds. The latter cannot be used for integrating QTLs with candidate genes.

QTLs for bud set have been determined in 2003 for two POP1 and POP2 pedigrees. Phenotypic data generated prior to the project were available for POP3a and. A total of 12 QTLs were found for bud set across three environments. Altogether, the genetic factors of bud set have been delineated to a 332 cM that equal about 12,8% of the genetic map-length. Expressed on physical sequence, this distance corresponds to about 58 Mb of the 308 Mb sequence currently contained in the linkage groups (18,8%). Importantly, two QTL regions were found in two different environments and two different breeding pedigrees (on linkage groups V and XI).

	cM	Mb
LGI	34	5,520
LGIII	46,3	8,400
LGIV	35	5,472
LGV	45/36,7	3,012
LGV		1,965
LGVI	13	2,538
LGVIII	20	0,436
LGIX	6,9	0,410
LGXI	40/12,3	4,680
LGXII	27	5,591
LGXIII	54	8,325
LGXVI	11,9	11,400

Table 4.4 All available bud set QTLs and their size on the physical sequence. Bud set was measured in autumn 2003, the year of establishment of the plantations.

The 955 considered candidate genes were equally distributed throughout the genome. 167 candidate genes co-localize with identified QTL regions for bud set on the genome, of these 31 have a regulatory function and 17 are transcription factors sensu stricto. Out of these genes, expressional candidate genes from linkage groups V and XI were chosen for SNP analysis in the association study below.

Nucleotide diversity and LD in the *P. nigra* population

In order to analyse sequence variation and linkage disequilibrium in *P. nigra*, 31 amplicons from 12 diverse poplar genotypes were amplified and sequenced directly from PCR products. PCR primers were designed to amplify 400-800 bp segment of each selected region and good sequences ranging from 200 to 600 bp were obtained from each segment (Table 4.5). Twenty amplicons were derived from BAC inserts of a poplar genomic library and chosen to represent pairs of single-copy sequences randomly selected in the poplar genome at an approximate distance of 100 kb from one another (average insert size of the library). Two of these sequences showed significant similarity

Number	Locus	Sequence length	Coding (exon) and noncoding (intron and BAC) regions			
1	D13Fend	315	noncoding			
2	D13Rend	247	noncoding			
3	C20Fend	190	noncoding			
4	C20Rend	190	1-87 intron	88-190 exon		
5	H11Fend	312	noncoding			
6	H11Rend	410	noncoding			
7	E01Fend	401	noncoding			
8	E01Rend	378	noncoding			
9	G19Fend	536	noncoding			
10	G19Rend	401	noncoding			
11	I13Fend	362	noncoding			
12	I13Rend	352	1-153 exon	154-352 intron		
13	A15Fend	367	noncoding			
14	A15Rend	271	noncoding			
15	M16Fend	315	noncoding			
16	M16Rend	326	exon			
17	O08Fend	302	noncoding			
18	O08Rend	349	noncoding			
19	PHYA100	572	exon			
20	PHYA102F	501	1-408 exon	409-501 intron		
21	PHYA102R	469	1-319 intron	320-469 exon		
22	PHYB1104	589	exon			
23	PHYB2108	545	exon			
24	PHYB2110	611	1-465 intron	466-611exon		
25	IAA2118	552	1-376 exon	377-465 intron	466-552 exon	
26	IAA2120	511	1-89 intron	90-314 exon	315-460 intron	461-511 exon
27	IAA2130	283	1-208 exon	209-283 intron		
28	CRY2124	588	1-407 exon	408-588 intron		
29	CRY1128	427	1-6 intron	7-413 exon	414-427 intron	

Table 4.5 loci analysed during the preliminary study

to plant genes when blasted against GenBank database, while no significant hits were found for the other 18 sequences. The strategy adopted to select the 20 BAC end amplicons resulted effective: the 576 BAC clones whose ends were sequenced (a total of 1152 sequences) provided 689 good-quality sequences after trimming and vector removal out of a total of 1152 sequences. Among these sequences 12 pairs of BAC ends satisfied the criteria detailed in Materials and Methods section. One pair of BAC ends was discarded because one sequence didn't give a single good quality amplification product, another pair of BAC ends was discarded after amplification because of sequencing problems from one of the sequences. So altogether in 10 out of 12 BAC clones both sequences worked well for both PCR amplification and direct sequencing. Eleven amplicons were derived from plant genes putatively involved in flowering time control [33]: indoleacetic acid-induced protein 2 (IAA2), phytochromes A, B1, B2, cryptochromes 1 and 2. A comparison between the sequences amplified in *P. nigra* and the related annotated gene sequences allowed the identification of protein coding portions in the sequenced amplicons, including in two BAC end sequences that putatively code for proteins. All sequences analysed were classified as introns, exons or genomic (if no similarity to protein sequences was found within the entire sequenced region) (Table 4.5).

The total amount of sequence analysed from the 31 regions was 13177 bp. We identified a total of 120 sequence variants from the sample of 24 chromosomes. Types and frequency of those variants are summarised in Table 4.6. 108 variants are single nucleotide changes (SNP) that occur on average every 122 bp; transitions were more frequent than transversions (ratio: 1.51). The frequency of substitutions in non-coding and genomic regions is almost twice higher than in coding sequences (one every 102 bp and one every 195 bp, respectively). The number of synonymous changes is twice that of non-synonymous ones (8 out of 23 changes resulted in amino acid substitution).

Twelve insertions/deletions (Indel) were detected in the 31 amplicons sequenced, with a frequency of 1 Indel every 1098 bp. Interestingly, only one Indel was found in 4497 bp of coding sequence, while no Indels were found in non-coding regions (1676 bp). All

remaining indels were found in the genomic anonymous BAC end sequences. With the exception of locus A15Rend where 4 Indels occur, the number of observed Indels per locus was 0 or 1. The single-nucleotide Indels are the most common: among the 12 Indels only one dinucleotide, one 4-bp and one 6-bp Indels were detected.

Parameter	Value	Comments
Number of loci screened	29	
Total length of amplicons, bp	11672	4753 coding, 6919 noncoding
Number of bases of sequence screened, bp	264856	
Number of all sequence variants (SNPs and Indels)	107	1 per 109 bp
Number of nucleotide substitutions (SNPs)	94	
Transitions / Transversions ratio	1.35	
Frequency of SNPs per bp	0.0081	1 per 124 bp
Frequency of SNPs per bp (coding)	0.0048	1 per 207 bp
Frequency of SNPs per bp (noncoding)	0.0103	1 per 97 bp
Synonymous changes / non-synonymous changes ratio	1.3	
Number of Indels	13	
Overall Indel frequency	0.0010	1 per 973 bp
Frequency of Indels per bp (coding)	0	
Frequency of Indels per bp (noncoding)	0.0017	1 per 576 bp
Total number of predicted haplotypes	112	
Mean number of predicted haplotypes per locus	4	
Mean nucleotide diversity (π)	0.0027	
Mean nucleotide diversity (synonymous sites)	0.0031	
Mean nucleotide diversity (non synonymous sites)	0.0009	
Mean nucleotide diversity (noncoding)	0.0012	

Table 4.6 Polymorphisms, haplotypes and nucleotide diversity (π) in the 29 loci analysed. Ch.: chromosome; [a] Triallelic InDel; [b] Sizes of the three InDels: one 6-nt deletion, one 6-nt insertion, one 11-nt InDel.

Information for each single locus concerning SNPs and Indels number, type and location is detailed in Table 4.7. No SNP was detected in 5 of the 31 loci. In 4 of these, however, an indel was observed, so only one locus did not show any sequence variation at all. SNPs and indel organisation into sequence haplotypes was also evaluated. Since black poplar is an outcrossing species, SNPs are frequently found in heterozygous condition within each individual. When more than one SNP is heterozygous within a locus then the

haplotype cannot be determined from the sequence. Haplotypes were then inferred by using the software Haplotyper (see Material and Methods) on the basis of a Bayesian algorithm. In addition to the locus that did not show any variant 9 other loci out of 31 showed a single variant (either SNP or indel) and thus required no haplotype inference (since the 2 haplotypes could be directly recognised). In the remaining 21 loci the sequence diversity is organised into a number of haplotypes ranging from 3 to 11 (Table 4.7) with an average of 5, where the most common haplotype is usually present in at least 40% of the chromosomes examined.

Locus	Ch. analysed	SNPs	SNP frequency (over 100 bp)	InDels	Predicted haplotypes	π (x1000)
D13Fend	22	1	0.32	0	2	0.99
D13Rend	24	0	0	1 (1 nt)	2	0
C20Fend	24	10	5.26	1 (2 nt)	10	20.50
C20Rend	22	2	1.05	1 (2 nt)	6	2.11
H11Fend	18	4	1.28	2 (3 nt)	6	4.06
H11Rend	22	8	1.95	0	6	8.37
E01Fend	20	2	0.50	0	3	1.55
E01Rend	18	2	0.53	1 (1 nt)	4	1.63
G19Fend	20	6	1.12	0	5	2.69
G19Rend	20	0	0	1 (1 nt)	2	0
I13Fend	18	2	0.55	1 (1 nt)[a]	6	2.46
I13Rend	20	3	0.85	0	3	2.80
A15Fend	24	3	0.82	1 (2 nt)	4	2.91
A15Rend	24	9	3.32	3[b]	9	9.51
M16Fend	22	1	0.32	0	2	0.78
M16Rend	24	1	0.31	0	2	0.26
O08Fend	24	1	0.33	0	2	1.30
O08Rend	24	0	0	1 (1 nt)	2	0
PHYA100	24	5	0.87	0	4	0.99
PHYA102F	24	2	0.40	0	3	0.74
PHYA102R	22	6	1.28	0	3	1.63
PHYB1104	24	3	0.51	0	4	1.24
PHYB2108	24	2	0.36	0	3	1.58
PHYB2110	24	10	1.63	0	6	5.60
IAA2118	24	2	0.36	0	3	0.33
IAA2120	24	6	1.17	0	4	1.26
IAA2130	22	2	0.71	0	3	1.71
CRY2124	24	0	0	0	1	0
CRY1128	24	1	0.23	0	2	0.68

Table 4.7 Polymorphisms and nucleotide diversity in the loci analysed

The average total nucleotide diversity (p) is 0.0024 (Table 4.7): it is higher for the synonymous sites and genomic and non-coding regions (0.0029 and 0.0022, respectively), while is lower for the non-synonymous sites (0.0009). Details for p values in each locus are shown in Table 4.8. To evaluate the allele distribution in the loci analysed and its possible deviations from neutral expectations, Tajima D statistics [34]

were computed, taking into account only SNPs. Negative values indicate an excess of low frequency alleles relative to neutral mutation-drift equilibrium, whereas positive values indicate a deficit of low frequency alleles relative to the expectation. Tajima D resulted on average positive (0.7299, Table 4.8). Only the two extreme values are statistically significant (Table 5), suggesting a departure from neutrality expectation for the two loci H11Rend (Tajima D= 2.14) and IAA2120 (Tajima D= -1.81). On the other

Region	π (x1000)	S.D. π (x1000)	π syn (x1000)	π non syn (x1000)	π non CDS (x1000)	Tajima's D
D13Fend	0.99	0.34	n.a.	n.a.	n.a.	0.24 (P > 0.10)
D13Rend	0	0	n.a.	n.a.	n.a.	n.d.
G14Fend	7.57	0.66	n.a.	n.a.	n.a.	1.38 (P > 0.10)
G14Rend	0	0	n.a.	n.a.	n.a.	n.d.
H11Fend	4.76	0.49	n.a.	n.a.	n.a.	0.92 (P > 0.10)
H11Rend	7.56	0.52	n.a.	n.a.	n.a.	2.14 (P < 0.05)
E01Fend	1.55	0.23	n.a.	n.a.	n.a.	0.24 (P > 0.10)
E01Rend	1.53	0.28	n.a.	n.a.	n.a.	0.06 P > 0.10)
G19Fend	2.46	0.82	n.a.	n.a.	n.a.	-0.69 (P > 0.10)
G19Rend	0	0	n.a.	n.a.	n.a.	n.d.
I13Fend	2.81	0.34	n.a.	n.a.	n.a.	1.22 (P > 0.10)
I13Rend	2.80	0.52	0	3.79	2.72	0.44 (P > 0.10)
N24Fend	5.43	0.78	n.a.	n.a.	n.a.	0,69 (P > 0.10)
N24Rend	2.32	0.43	n.a.	n.a.	n.a.	0.44 (P > 0.10)
A15Fend	3.03	0.37	n.a.	n.a.	n.a.	0.97 (P > 0.10)
A15Rend	10.64	2.36	n.a.	n.a.	n.a.	-0.42 (P > 0.10)
M16Fend	0.78	0.34	n.a.	n.a.	n.a.	- 0.17 (P > 0.10)
M16Rend	0.26	0.23	0	0.33	0	-1.16 (P > 0.10)
O08Fend	1.00	0.29	n.a.	n.a.	n.a.	0.48 (P > 0.10)
O08Rend	0	0	n.a.	n.a.	n.a.	n.d.
PHYA100	0.99	0.49	1.78	0.75	0	-1.67 (0.10 > P > 0.05)
PHYA102F	2.68	0.52	0	3.47	2.69	-0.50 (P > 0.10)
PHYA102R	1.63	0.87	2.74	0	2.12	-1.66 (0.10 > P > 0.05)
PHYB1104	1.27	0.24	4.80	0.19	0	-0.17 (P > 0.10)
PHYB2108	1.58	0.18	6.78	0	0	1.34 (P > 0.10)
PHYB2110	5.65	0.58	13.19	0	6.43	0.96 (P > 0.10)
IAA2118	0.33	0.20	0.83	0	1.02	-1.51 (P > 0.10)
IAA2120	1.26	0.72	4.86	0	1.42	-1.81 (P < 0.05)
IAA2130	1.71	0.48	0	3.06	0	-0.26 (P > 0.10)
CRY2124	0	0	0	0	0	n.d.
CRY1128	0.68	0.24	3.20	0	0	0.14 (P > 0.10)

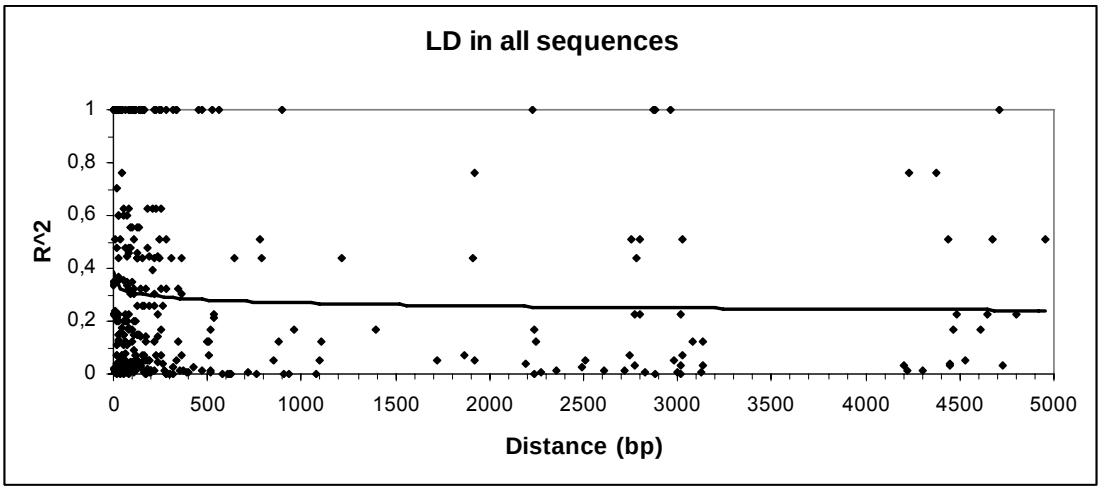
Table 4.8 Tajima D statistic for loci of interest and analysis of nucleotide polymorphism frequency

significance for the D values at a significance level of 0.05 and thus expected 1.3 significant tests even in absence of any real deviation from the mutation-drift

equilibrium. In order to evaluate the levels of linkage disequilibrium (LD) within the sequenced regions, LD measures D' and r2 were calculated for SNP loci within each BAC end sequence and each gene (by combining the information from the different fragments of the same gene). Only the composite plot of r2 values of LD are presented (Figure 4.3). LD over longer physical distances (approximately 100 kb) was also evaluated by computing r2 values between pairs of BAC ends belonging to the same physical clone.

As a control for the background levels of LD in the black poplar genome (i.e. empirical null distribution) r2 values were also computed between all BAC ends excluding the pairs deriving from the same clone: since BAC clones were chosen at random these should represent physically unlinked regions of the genome (Figure 4.4). Low r2 values were found (average = 0.3) within the single regions but no significant decline could be observed within 5 kb in the loci (all corresponding to candidate genes) where we had SNPs spanning such distances. When r2 was computed across regions that are on average at 100 kb distance (BAC ends in the same clone) it decreases considerably (average r2=0.09) and is very similar to the level of LD observed between unlinked loci (average r2=0.09). The Mann-Whitney U test shows that the distribution of LD estimates as measured by r2 for the short range distances (0-5000 bp) is significantly different from the distribution for long range distances (100 kb) ($P<0.001$) as well as from that for unlinked regions ($P<0.001$) but that no significant difference is detected between the latter two distributions ($P=0.37$).

Figure 4.3 R^2 measures for linkage disequilibrium in all loci analysed. A logarithmic trend line is included in the plot.



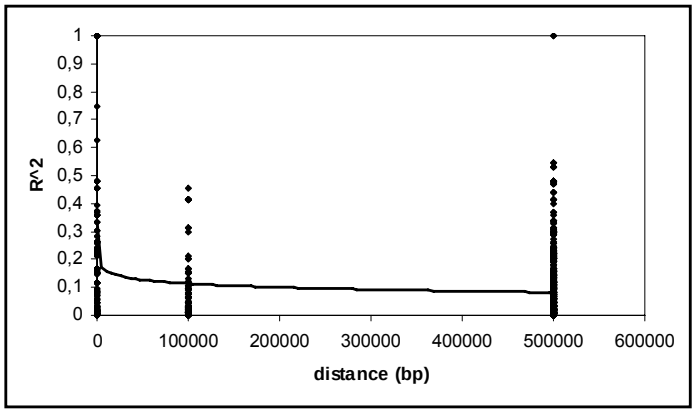


Figure 4.4 R^2 measures for linkage disequilibrium over long physical distances; SNPs within, between and among BAC ends are considered. A logarithmic trend line is included in the plot.

Linkage disequilibrium was also calculated in the two pools of genomic and coding sequences (data not shown). The r2 values were only considered between SNPs in the limited range of 300 bp in order to produce two homogeneous and comparable samples from the point of view of physical distances between SNPs. LD results higher (average

r2=0.33) in the candidate gene sequences than in non-coding genomic sequences (BAC ends) (average r2=0.27) but the two distributions of r2 values are not significantly different when tested with the Mann-Whitney test (P= 0.24).

Association analysis of bud-set within the *P. nigra* population

Nine candidate genes (Table 4.4) out of the 37 preliminary analysis were selected, focusing on those phytochromes, cryptochromes, signal integrators, vernalization factors and circadian clock factors suggested as more interesting by the recent literature and by the current researches on QTL and transcriptome profiling for bud set presented in the results section. A total of 27 informative SNPs (TagSNP in the Table 4.4) were successfully genotyped; the individuals used for the tagSNP discovery were also included in each genotyping experiment as positive controls. The genotypic data obtained were preliminary used to compute the F-statistics (estimated as average of Fst of each SNP considered as an independent locus) using the software Arlequin 2.0 [35]. Low-to-moderate levels of genetic differentiation (Fst from 0.05 to 0.1; total Fst= 0.06) were observed in the candidate genes analysed (Figure 4.5), even if the distribution of Fst values for each polymorphic locus may suggest the presence of few outlying SNPs (Fst > 0.15).

The SNP genotypic data of the candidate genes were then used to perform association analysis, using the software TASSEL (Trait Analysis by Association, Evolution and Linkage; http://www.maizegenetics.net/tassel) using the phenotypic data for bud set. Among the different bud set traits measured, P9 considered 13 traits suggested by P2 and P3 (they fitted the raw data) as the most representative of the phenotypic variation: 5 traits express the onset date of each bud set score from 2.5 to 0.5 with a 0.5 increment, expressed in cumulative night length (cnl); 3 traits represent the area under the bud set curve (AUBSC) of the whole process (score 3 to 0) and of the two subprocesses (sub1: score 3 to 1.5; sub2: score 1.5 to 0); 2 traits express the duration of defined subprocesses (adjusted mean values provided by P2); 2 traits represent important parameters for practical application in tree breeding (adjusted mean values provided by P2); 1 trait

express again the date of onset of score 2.5 provided as adjusted mean data. The association was computed by a least square analysis according to the General Linear Model tool provided by the software, where the SNP genotypic data were considered as the only effect on the model and the significance of marker effect was expressed using 1000 permutations. At the moment no structure and kinship data were included in the model as not still available.

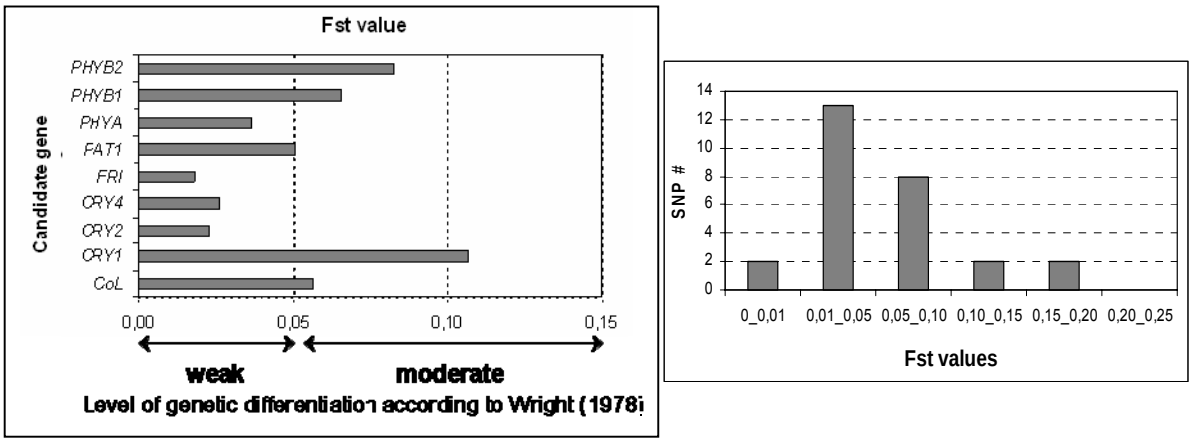


Figure 4.5 Genetic differentiation in the candidate genes analysed, as computed by Arlequin 2.0. In the left, Fst values computed within each candidate gene. In the right, classes of Fst values of all SNPs genotyped.

As shown in Figure 4.6, only few polymorphic sites within two different candidate genes (*Cry4* and *CoL*) provided significant associations; it's interesting to stress that both candidates showed low Fst values (see Figure 4.5). *Cry4* is one of the 4-member gene family of *cryptochromes* found in the poplar genome and has been annotated on the linkage group V of the *P. trichocarpa* genome. The tagSNP identified within this gene corresponds to a non coding transversion (G/T) in the first intron. It showed a significant association (p= 1,50E-10) to 8 out of 13 traits considered (six concerning the onset of the score and 2 concerning the important breeding traits), explaning 10% of the variation (Rsq= 0.11). *CoL* is one of the two members of the *Constans-like* genes annotated in the *P. trichocarpa* genome, but it has not been mapped into a linkage group, yet. Among the four tagSNPs of this gene, a transition A/G within the intron showed a significant association (p= 1,52E-05 and p= 1,0E-04) to the same 8 traits mentioned above. It

explanes 6 % of the variation (Rsq= 0.06) and has an additive effect on the phenotypes. Both SNPs don't correspond to those putative outlying loci identified in figure 4, making them attractive as potential causative polymorphisms for the bud set phenotypic variation. For both SNPs, the hypothesis of their involvement as source of alternative splicing in the genes concerned can be considered and checked by computational means.

Additionally to the two mentioned polymorphic sites, higher p values than the background can be observed in the surrounding genomic tracts of CoL's most significant SNP. This makes Constans-like genes very interesting to be further analysed as target genes in the growth cycle/dormancy pathway.

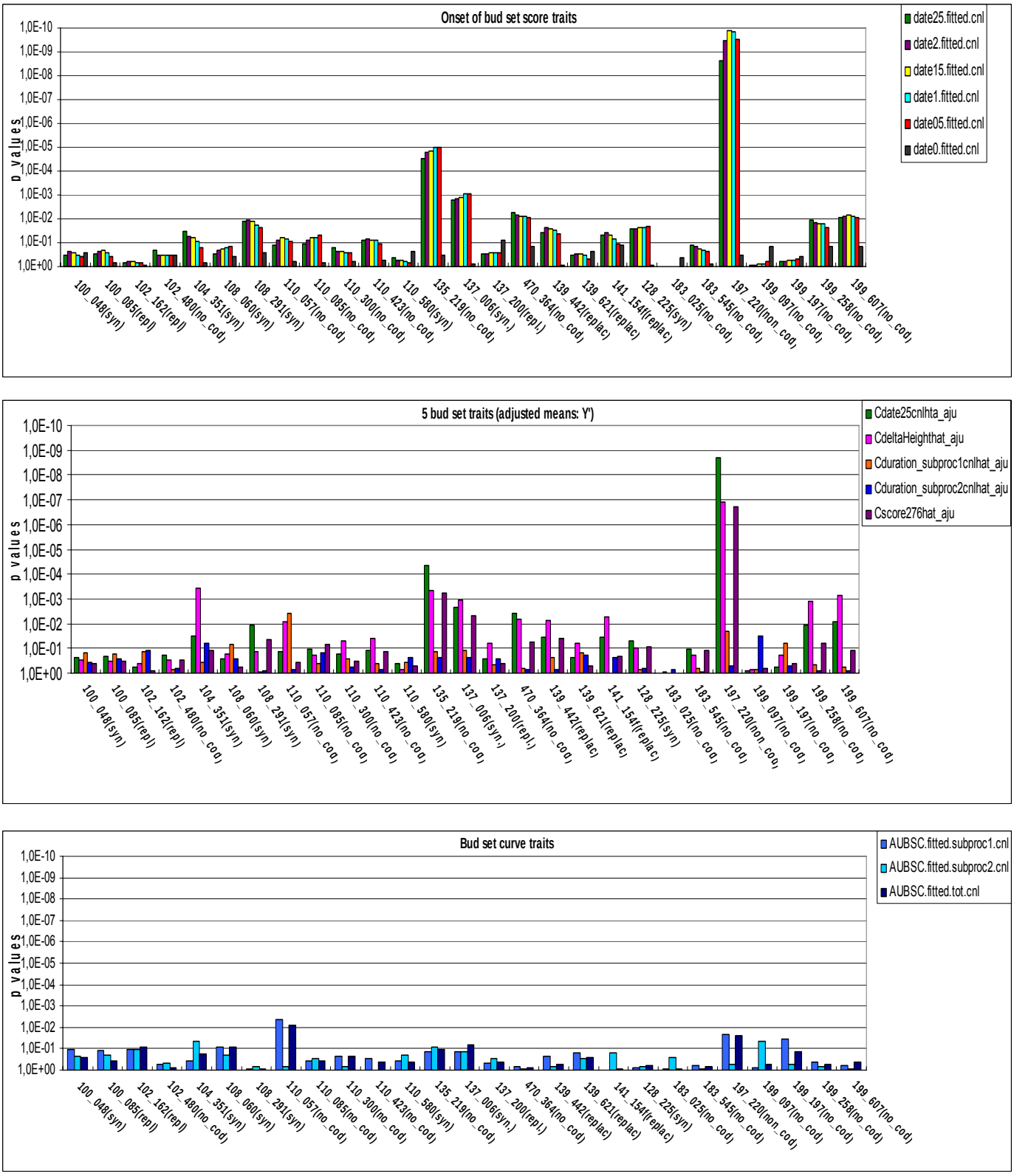


Figure 4.6 Plot of the P values obtained from the association analysis of 27 SNPs markers, genotyped in the LD collection (398 clones), and 13 bud set traits. Data were obtained from TASSEL.

4.4 Discussion

Molecular genetic analysis (using AFLPs, RFLPs, SSRs) has been employed in *Populus nigra*, one of the native European poplar species, to estimate genetic diversity and to assess the contribution of clonal and sexual reproduction to its populations. As in other forest trees, mainly because of the time-to-flower limitation and strong inbreeding depression, it is difficult to generate recombinant inbred lines or set up back-cross programs in order to study genetic variation and utilize it for mapping purposes [35], especially if high resolution mapping is required with the goal of positionally cloning genes [36]. A new approach that allows us to overcome these limitations in woody perennials utilises the presence of linkage disequilibrium to perform association mapping [37]. The association mapping approach uses natural populations, where genotypes differing for the phenotype of interest are chosen and then genotyped to identify polymorphisms associated with the trait/phenotype. The number and density of markers to be employed in such an approach, called whole genome scan, will vary depending on the extent of linkage disequilibrium present in the individuals being analysed. As an alternative, the individuals being phenotyped can be analysed only for a limited number of genes that have been chosen as candidates based on previous knowledge on their possible biochemical or metabolic role in connection to the trait being studied, in the so called candidate gene approach. This was the approach that was eventually chosen here, to study the link between an important phonological trait and the underlying genes determining that trait.

Our study on nucleotide diversity is based on 12 *Populus nigra* trees collected from the natural population in different regions across France (10 individuals) and Italy (2 individuals), apparently unaffected by human selection. Since black poplar populations are usually in Hardy Weinberg equilibrium [19], this corresponds to sampling genetic variation in 24 independent chromosomes. We analysed 31 regions of the poplar genome. The availability of the genome sequence of *P. trichocarpa* and of a genomic BAC library of *P. nigra* allowed us to select 20 single-copy genomic regions from 10 different BAC clones and 11 regions from 6 genes for sequence analysis. The selection of the single

copy genomic regions from BAC clones proved to be effective since single and good-quality PCR products were obtained to be used directly in the sequencing reaction. Variation analysis was performed through the direct sequencing of PCR products method, that has demonstrated to be a high-throughput, relatively rapid and low-cost way for discovering SNPs [2]. Black poplar is an outcrossing species with individuals in Hardy Weinberg equilibrium. Many of the SNPs within each individual will therefore be in heterozygous condition. This complicates two phases of the SNP analysis: SNP identification first, since polymorphisms are not only found as differences between different sequences but also as heterozygous positions within a sequence, i.e. positions with two overlapping signals, and the haplotype derivation later, since the phase of the adjacent SNP alleles is not known. The use of PolyPhred software [26] helped greatly to identify heterozygous positions in a semiautomated way and the use of the Haplotyper software [26] allowed the inference of intralocus haplotypes. The overall frequency of SNPs (1/122 bp) found in the 31 loci is almost one half the overall frequency found in crop species (Zea mays 1/61 bp, Ching et al. 2002; *Oryza sativa* 1/89 bp, [38]; *Vitis vinifera* 1/65 bp, G. Prete & M. Morgante, unpublished data) and in the forest tree *Picea abies* (1/75 bp; S. degli Ivanissevich & M. Morgante, unpublished data). Indels seem to be rare both in coding and in non coding regions, with an overall frequency of one every 1.1 kb, as opposed to one every 0.1 kb in maize [5]. As expected both SNP as well as Indel frequencies increased in non coding and random genomic regions as opposed to coding ones. The frequency data suggest a low overall nucleotide polymorphism but still sufficient to give at least one sequence polymorphisms (either SNP or Indel) in 30 out of 31 regions considered. It therefore appears that sequence polymorphisms wouldn't be a limiting factor for the development of association mapping strategies in black poplar. The low sequence variability is even more evident if nucleotide diversity (p) estimates are considered. The overall p in the 31 loci examined is 0.0024 that is lower than the averages found in maize (p= 0.0063, [5] sugar beet (p= 0.0076, Schneider et al., 2001), Norway spruce (p= 0.0043, S. degli Ivanissevich & M. Morgante, unpublished data), grape (p= 0.0056, G. Prete & M. Morgante, unpublished data), and higher only than that found in mammals (human: p= 0.0004/0.0008, Cargill et al., 1999; mouse: p=0.001, [40]). The low variability is also supported if synonymous and non-synonymous

nucleotide diversities are compared to the related values found in other forest trees. Synonymous p of poplar (0.0029) is 66% of that found in silver birch [14] and pine [12], while the non-synonymous one is very low in all the three species (0.0009, 0.000052 and 0.0004, respectively). One reason for the low SNP frequency could be the small sample of chromosomes analysed (24) that caused us to miss rare SNP polymorphisms [41].

The only Europe-wide study of genetic diversity in the species using three different marker systems shows moderate levels of differentiation between populations, with an FST estimate for codominant markers of 0.12 [19]. The same study shows that France and Italy are among the most diverse regions and that very few alleles unique to specific regions are found. The results on low nucleotide variability should therefore not have been strongly biased downwards by a sampling of individuals that did not encompass the whole distribution range of the species. Our data on low nucleotide variability are not consistent with genetic variation data on black poplar populations based on isozymes or microsatellites, that suggest an high overall diversity level (average gene diversity HS for microsatellites: 0.76; average gene diversity HS for isozymes: 0.13; Storme et al., 2004). In other studies in *Populus nigra* high levels of diversity were also detected in populations within a river system and differentiation was mostly found between populations from different river systems. Our results that show low nucleotide diversity could be explained by either a small effective population size or a low nucleotide substitution rate or by directional selection in the loci considered that may be not fully neutral. We also observed a relatively high number of haplotypes, with an average of 5 haplotypes for the 21 loci that had at least two mutations (SNPs and/or Indels). As a comparison, in a sample of 38 maize chromosomes 3.4 haplotypes were observed on average, even though the SNP frequency (1 per 61 bp) and the nucleotide diversity (0.0063) were both much higher than the ones we observed in poplar.

It is important to evaluate the structure and the extent of linkage disequilibrium in the genome of interest before undertaking an association study. We used the r2 measure of LD rather than the D' one because of the relatively small sample size we used that makes D' very often assume absolute values of 1 [4]. Even though the absolute values of D' that

we observed are much higher, the trends with respect to physical distances are very similar to those observed for r2 (data not shown). LD within each single sequenced region was not very high, with average r2 values of 0.3 at 100 bp distances. When we calculate LD over the length of the genes we analysed by comparing the different regions within each gene, the decay of LD doesn't seem to be dramatic over distances of up to 5 kb, even though the sample size is much smaller for the longer distances than the smaller ones. Selection for adaptive traits may create linkage disequilibrium among chromosomal regions containing major genes for the selected traits [8]. Consistent with the lack of selection in the loci analysed in the present study, LD resulted to be not significantly different between coding and non-coding (genomic) regions. We used SNPs from BAC end sequences from the same BAC clone to analyse LD over longer distances, i.e. over distances around 100 kb. LD values do not seem to follow a normal distribution when averaged over loci since they can vary widely between different genomic regions. In order to assess the significance of the levels of LD observed across specific physical distances we employed a non parametric test to compare these distributions of r2 values against the one obtained from physically unlinked SNPs and assumed this latter one to represent an empirical null distribution or in other terms to represent background LD noise (which may include spurious LD as well). The test shows that while the LD levels within each BAC end sequence are significantly different (higher) from the background noise, the levels observed at 100 kb distances are not different from those observed between physically unlinked loci. On the basis of our data, in *P. nigra* there is evidence for a low but significant LD that may extend to over 5000 bp but definitely declines within 100 kb. The observation of low levels of LD in poplar even within relatively short distances is consistent with the finding of a relatively high number of haplotypes within each sequenced region. Very few and limited data sets exist for LD levels in plants. In maize LD structure depends on the population under analysis, with elite inbred lines showing strong LD and less selected material harboring a rapid decay. In Scots pine very low levels of linkage disequilibrium have been detected within the PAL gene [12], and in Norway spruce LD has also been found to decay within a few hundred bases [8]. These observations have implications for the choice of strategies for association mapping studies: SNP frequency seems to be sufficient even within genic regions to allow for the

use of genes as sources of SNPs for such studies. The fact that linkage disequilibrium does not seem to extend over large regions and appears to be negligible at 100 kb distances is in favor of the use of a candidate gene approach where large LD blocks could cause the appearance of false associations that are not due to the involvement of the gene being tested but of one of the genes nearby that is in disequilibrium with it. It therefore appears that candidate gene association mapping approaches are a definite possibility in black poplar.

The availability of all genes within the *Populus trichocarp*a draft genomic sequence and the possibility to successfully amplify them in other poplar species makes this an especially attractive option for the identification of genes involved in a variety of traits of interest where considerable information exists on genes involved in their determination from model species. The promising results, obtained with the present association study, leads to extend the analysis to the whole genomic region containing the two genes which gave positive association and to look for the putative causative mutation which can explain the adaptation of the trait. Moreover, the association study can be extended to a wider panel of candidate genes (different from those suggested by the literature) taking advantage on the final findings on bud set candidates made by the other partners who used other genomic approaches, including transcriptome profiling.

4.5 Conclusions

We have undertaken novel analysis of linkage disequilibrium in *Populus nigra*, an important native European tree species. We have quantified nucleotide diversity and estimated the amount of LD in the population. Following this and taking a candidate gene approach, we have identified positive associations between the phonological traits of bud set to identify associations between budset and important genes though to control flowering time in Arabidopsis.

4.6 References

1. Buckler IV ES, Thornsberry JM (2002) Plant molecular diversity and applications to genomics. *Current Opinion in Plant Biology,* 5, 107-111

2. Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. *Current Opinion in Plant Biology,* 5, 94-100

3. Pflieger S, Lefebvre V, Causse M (2001) The candidate gene approach in plant genetics: a review. *Molecular Breeding*, 7, 275-291

4. Flint-Garcia SA, Thornsberry JM, Buckler ES 4th (2003) Structure of linkage disequilibrium in plants. *Annual Review of Plant Biology*, 54, 357-374

5. Ching A, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey S, Morgante M, Rafalski AJ (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. *BMC Genetics*, 3, 19-32

6. Wang RL, Stec A, Hey J, Lukens L, Doebley J (1999) The limits of selection during maize domestication. *Nature*, 398, 236-239

7. Hilton H, Gaut BS (1998) Speciation and domestication in the maize and its wild relatives: evidence from the globulin-I gene. Genetics 150, 863-872.

8. Rafalski A, Morgante M (2004) Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. *Trends in Genetics*, 20, 103-111

9. Nordborg M, Borevitz JO, Bergelson J, Berry CC, Chory J, Hagenblad J, Kreitman M, Maloof JN, Noyes T, Oefner PJ, Stahl EA, Weigel D (2002) The extent of linkage disequilibrium in Arabidopsis thaliana. *Nature Genetics*, 30, 190-193

10. Le Corre V, Roux F, Reboud X (2002) DNA polymorphism at the FRIGIDA gene in Arabidopsis thaliana: extensive nonsynonymous variation is consistent with local selection for flowering time. *Molecular Biology and Evolution*, 19, 1261-1271

11. Kawabe A, Yamane K, Miyashita NT (2000) DNA polymorphism at the cytosolic phosphoglucose isomerase (PgiC) locus of the wild plant Arabidopsis thaliana. *Genetics*, 156, 1339-1347

12. Dvornyk V, Sirvio A, Mikkonen M, Savolainen O (2002) Low nucleotide diversity at the pal1 locus in the widely distributed Pinus sylvestris. *Molecular Biology and Evolution*, 19, 179-188

13. Garcia-Gil MR, Mikkonen M, Savolainen O (2003) Nucleotide diversity at two phytochrome loci along a latitudinal cline in Pinus sylvestris. *Molecular Ecology*, 12, 1195-1206

14. Jarvinen P, Lemmetyinen j, Savolainen O, Sopanen T (2003) DNA sequence variation in BpMADS2 gene in two populations of Betula pendula. *Molecular Ecology*, 12, 369-384

15. Legionnet A, Lefevre F (1996) Genetic variation of the riparian pioneer tree species Populus nigra L. I. Study of population structure based on isozymes. *Heredity*, 77, 629-637

16. Arens P, Coops H, Jansen J, Vosman B (1998) Molecular genetic analysis of black poplar (Populus nigra L.) along Dutch rivers. *Molecular Ecology*, 7, 11-18

17. Winfield MO, Arnold GM, Cooper F, Le Ray M, White J, Karp A, Edwards KJ (1998) A study of genetic diversity in Populus nigra subsp. betulifolia in the Upper Severn area of the UK using AFLP markers. *Molecular Ecology*, 7, 3-10

18. Imbert E, Lefevre F (2003) Dispersal and gene flow of Populus nigra (Salicaceae) along a dynamic river system. *Journal of Ecology*, 91, 447-456

19. Storme V, Vanden Broeck A, Ivens B, Halfmaerten D, Van Slycken J, Castiglione S, Grassi F, Fossati T, Cottrell JE, Tabbener HE, Lefevre F, Saintagne C, Fluch S, Krystufek V, Burg K, Bordacs S, Borovics A, Gebhardt K, Vornam B, Pohl A, Alba N, Agundez D, Maestro C, Notivol E, Bovenschen J, van Dam BC, van der Schoot J, Vosman B, Boerjan W, Smulders MJ (2004) Ex-situ conservation of Black poplar in Europe: genetic diversity in nine gene bank collections and their value for nature development. *Theoretical and Applied Genetics*, 108, 969-981

20. Ingvarsson PK, García MV, Hall D, Luquez V, Jansson S (2006) Clinal Variation in *phyB2*, a Candidate Gene for Day-Length-Induced Growth Cessation and Bud Set, Across a Latitudinal Gradient in European Aspen (*Populus tremula*). *Genetics*, 172, 1845-1853

21. Neale D. and Savolainen O (2004) Association genetics of complex traits in conifers. *Trends in Plant Science*, 9, 325-330

22. Pichot C, Ducros ET (1989) estimation of genetic-parameters in eastern cottonwood (populus-deltoides bartr) - consequence for the breeding strategy. *Annales des Sciences Forestieres*, 46, 307-324

23. Breyne P, Zabeau M (2001) Genome-wide expression analysis of plant cell cycle modulated genes. *Current Opinion in Plant Biology,* 4, 136-142

24. Huang X, Madan A (1999) CAP3: A DNA Sequence Assembly Program. *Genome Research*, 9, 868-877

25. Chou HH, Holmes MH (2001) DNA sequence quality trimming and vector removal. *Bioinformatics*, 17, 1093-1104

26. Nickerson DA, Tobe VO, Taylor, SL (1997) PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Research*, 25, 2745-2751

27. Gordon D, Abajian C, Green P (1998) Consed: A graphical tool for sequence finishing. *Genome Research*, 8, 195-202

28. Gordon D, Desmarais C, Green P (2001) Automated Finishing with Autofinish. *Genome Research*, 11, 614-625

29. Niu T, Struk B, Lindpaintner K (2001) Statistical considerations for genome-wide scans: design and application of a novel software package polymorphism. *Human Heredity,* 52, 102-109

30. Rozas J, Rozas R (1999) DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics*, 15, 174-175

31. Ding K, Zhou K, He F, Shen Y (2003) LDA - A java-based linkage disequilibrium analyzer. *Bioinformatic,*. 19, 2147-2148

32. Schneider K, Weisshaar B, Borchardt DC, Salamini F (2001) SNP frequency and allelic haplotype structure of Beta vulgaris expressed genes. *Molecular Breeding,* 8, 63-74

33. Simpson GG, Dean C (2002) Arabidopsis, the Rosetta stone of flowering time? *Scienc,*. 12, 285-289

34. Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123, 585-595

35. Taylor G (2002) Populus: Arabidopsis for forestry. Do we need a model tree? *Annals of Botany*, 90, 681-689

36. Morgante M, Salamini F (2003) From plant genomics to breeding practice. *Current Opinion in Biotechnology*, 14, 214-219

37. Neale D. and Savolainen O (2004) Association genetics of complex traits in conifers. *Trends in Plant Science*, 9, 325-330

38. Han B, Xue Y (2003) Genome-wide intraspecific DNA sequence variations in rice. *Current Opinion in Plant Biology*, 6, 134-138

39. Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, Shaw N, Lane CR, Lim EP, Kalyanaraman N, Nemesh J, Ziaugra L, Friedland L, Rolfe A, Warrington J, Lipshutz R, Daley GQ, Lander ES (1999) Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nature Genetics*, 22, 231

40. Lindblad-Toh K, Winchester E, Daly MJ, Wang DG, Hirschhorn JN, Laviolette JP, Ardlie K, Reich DE, Robinson E, Sklar P, Shah N, Thomas D, Fan JB, Gingeras T, Warrington J, Patil N, Hudson TJ, Lander ES (2000) Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. *Nature Genetics*, 24, 381-386

41. Renwick A, Bonnen PE, Trikka D, Nelson DL, Chakraborty R, Kimmel M (2003) Sampling properties of estimators of nucleotide diversity at discovered SNP sites. *International Journal of Applied Mathematics and Computing Science*, 13, 385-394

42. Pluzhnikov A, Donnelly P (1996) Optimal sequencing strategies for surveying molecular genetic diversity. *Genetics*, 144, 1247-1262

5. *Populus* genomics and genetical genomics to identify tree genes for disease resistance and response to environment

W Boerjan, S Jansson, G Taylor, F Martin, NR Street, AM Rae, J Tucker, A. Rohde, V Jorge, V Storme, M Steenackers, S Jansson, O Skögstrom, A Sjodin , S Duplessis, V Jorge, C Bastien, C Rinaldi, A Kohler

5.1 Introduction

The final aim of QTL mapping is to identify the causal gene responsible for the trait. Positional cloning is difficult to achieve in out-breeding species that suffer inbreeding depression, have long generation times and large genomes. Currently, there are no examples of genes that have been positionally cloned from any forest tree, even for traits showing Mendelian segregation, such as resistance to *Melampsora* spp.

Several strategies exist to identify the genes underlying QTLs. One is the candidate gene approach in which genes assumed to be involved in the trait are genetically mapped to see whether their map positions coincide with QTLs mapped for that trait. However, the large linkage generated in artificial crosses does not allow making firm conclusions on the role of the candidate gene in determining the trait.

A second strategy combines a QTL analysis of phenotypes with a QTL analysis of gene expression levels (eQTL) in a mapping pedigree, a strategy called genetical genomics. The first such approach in forest trees has been carried out in an interspecific *Eucalyptus* backcross population, where QTLs for diameter growth co-localized with eQTLs for lignin-related genes, suggesting that growth and lignin characteristics are controlled by the same loci [1]. None of the lignin-related genes themselves mapped at the growth QTL, except for the gene encoding S-adenosyl-L-methionine synthetase (SAMS). Because down-regulation of SAMS in maize results in reduced lignin content, SAMS is a good

candidate regulator of both lignin and growth in eucalyptus. A similar strategy to identify candidate genes for yield was followed at the proteome level in maritime pine, where a QTL for biomass production co-localizes with the *GS* gene and a protein quantity QTL (PQTL) that controls the abundance of GS [2]. Still, these candidates genes may remain associated with the trait merely because of linkage in the mapping pedigree. The lack of resolution in mapping candidate genes and QTL alleles can then be overcome by association genetics, using natural populations in which the long evolutionary history has broken up the linkage between markers and genes.

Within POPYOMICS, the genetical genomics approach has been tested on poplar for three different traits, tolerance to drought stress, tolerance to ozone stress and quantitative resistance to infection with the fungus *Melampsora larici-populina*. The genetical genomics approach used was slightly different for the different traits, and is explained below. For each trait, differentially expressed candidate genes have been identified, which will now need further support with association genetics [3].

5.2 Genomic Methods Employed Within POPYOMICS

The *Populus* genome sequence

Background

When the POPYOMICS proposal was written, we did not dare to hope that the full genome sequence for *Populus* would be available during the timeframe of the project. However, when in 2002 an International consortium set out to sequence the *Populus trichocarpa* genome, new avenues were opened up not only for our project but also for the *Populus* community in general, and resulted in substantial reprioritization within POPYOMICS. Several of the Popyomics partners (Partner Umeå, Nancy, Gent) got involved in the effort, and, shot-gun sequencing of the genome to about 8x depth was completed, financed by the US Department of Energy (DOE) and coordinated by the Oak Ridge National Laboratory (ORNL) and Joint Genome Institute (JGI).

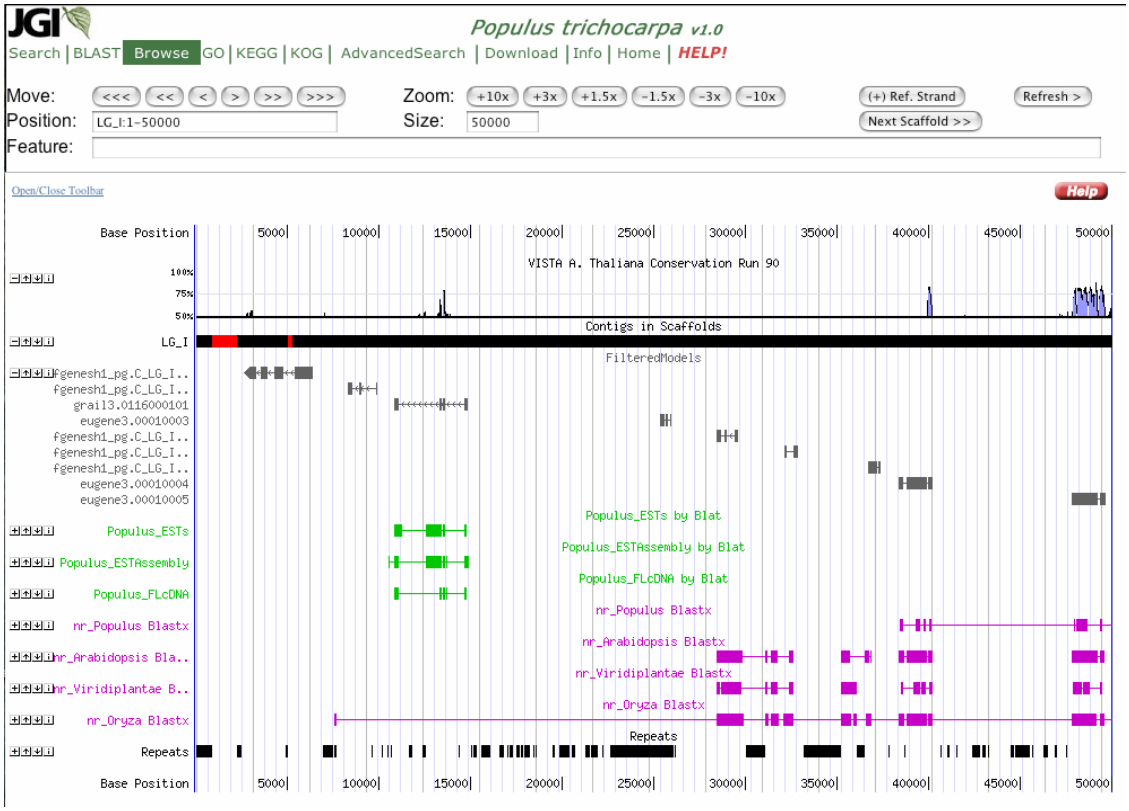


Plate 5.1 Screenshot of the public JGI *Populus* genome browser.

The annotated genome sequence

After the genome was assembled using the JAZZ algorithm, repetitive sequences were identified and mapped and the assembly has been masked for these regions. Genes were predicted, in non-masked regions, using 4 different gene calling algorithms, and a non-redundant set of about 45 000 gene models was created, representing the first iteration towards determining the gene number of Populus. These genes have been annotated and the database containing these data (www.jgi.doe.gov/poplar) was opened up to the scientific community on September 21 2004. A screenshot of the database is shown in Plate 5.1. About 3/4 of the euchromatin has been mapped to the linkage groups (chromosomes) and mapping process is ongoing. Therefore the chromosomal location of almost 75 % of the Populus genes has been determined.

The *Populus* genome turned out to be a recently duplicated genome (Plate 5.2) important for some of the results from our mapping activities.

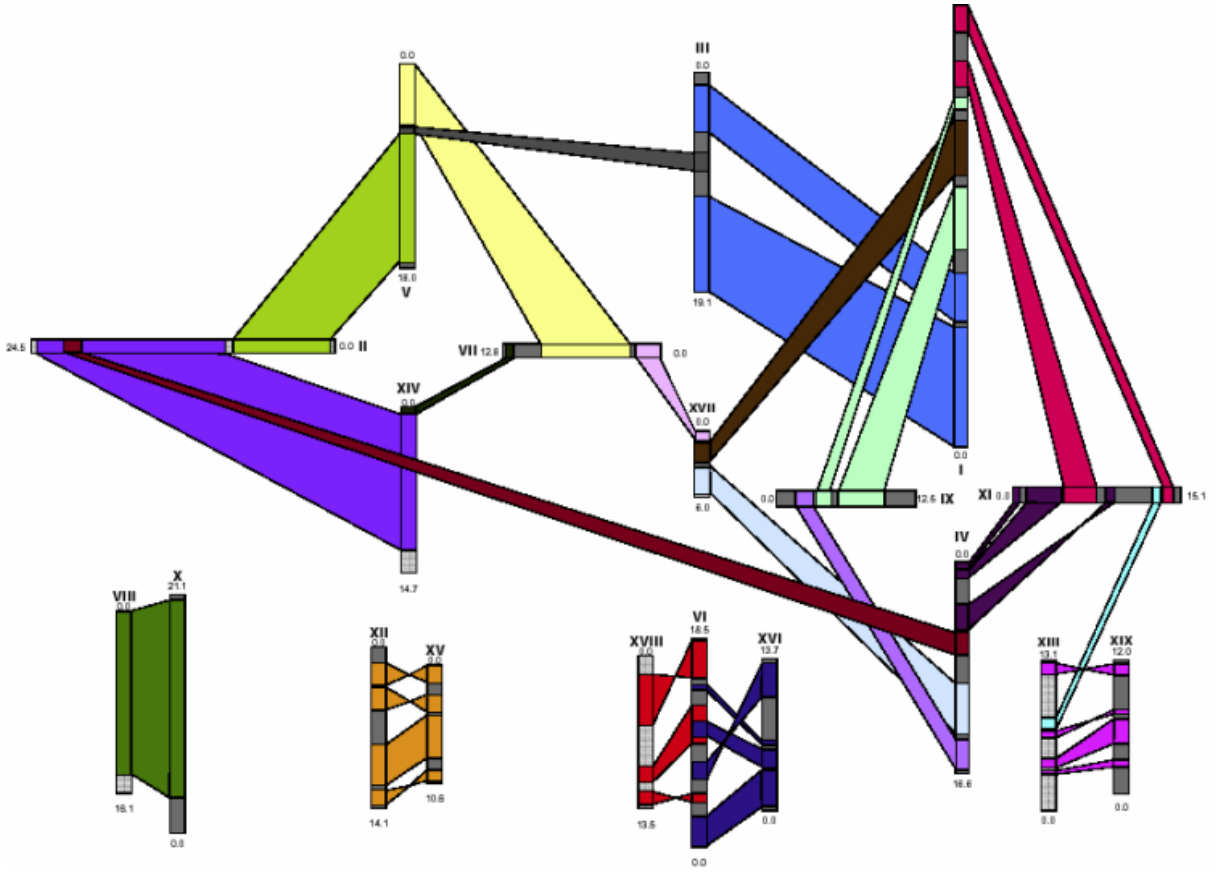


Plate 5.2 . The structure of the *Populus* genome from [7].

Transcriptomics

It is of very high importance to get information about the expression patterns of genes. Within POPYOMICS, we have used complementary techniques to get this information, including EST sequencing and cDNA microarrays.

Expression profiling – EST database

ESTs are short sequence tags of expressed genes, obtained by sequencing of randomly picked clones from a cDNA library prepared from mRNA sampled from a given tissue. Several of the Popyomics partners had performed extensive EST sequencing prior to the

POPYOMICS project, financed by other funds. However, within POPYOMICS the databases resulting from the EST efforts were made available to the partners – and the public – and considerable developments were made to facilitate data retrieval.

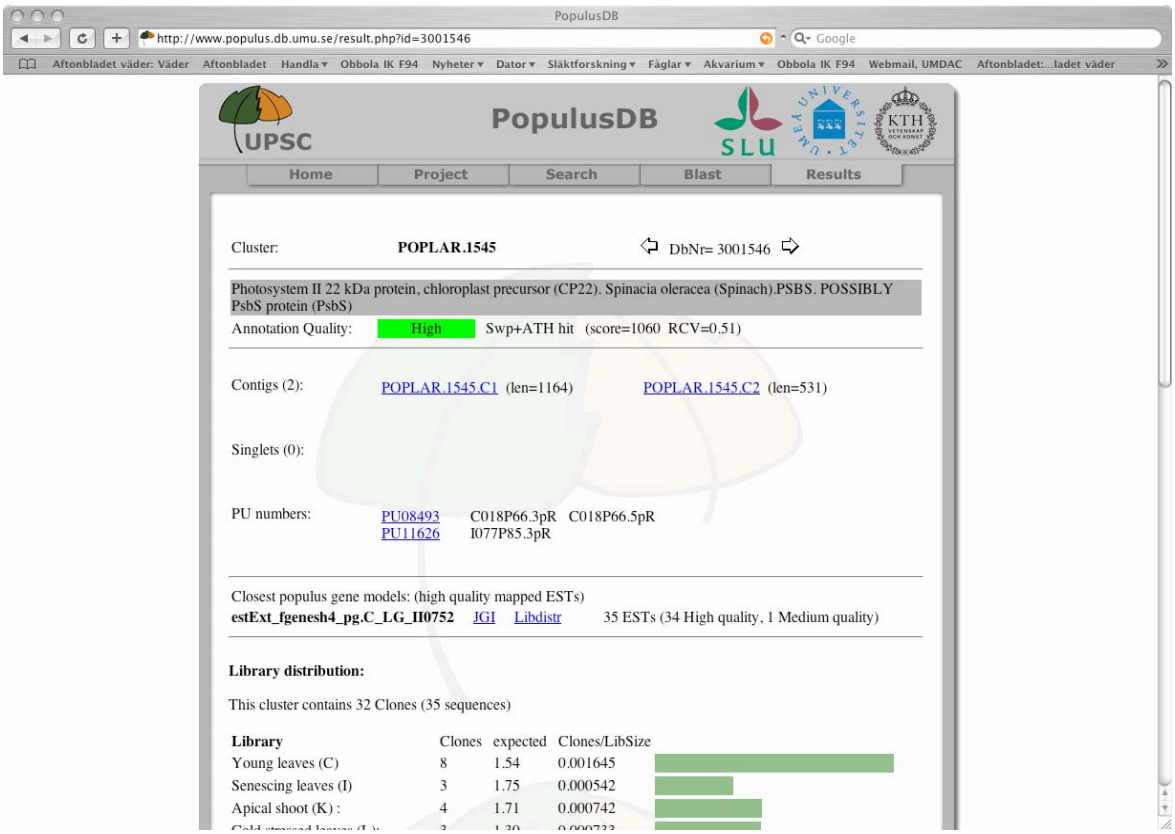


Plate 5.3 Screenshot of the public PopulusDB

The largest EST database has been created in the institute of partner 7, and this will in the following be briefly described. In general, the structure of the work in the institute of the other partners was similar. PopulusDB (www.populus.db.umu.se) can, among other things, be used to on the fly get information about in which *Populus* tissue, and sometimes also under which conditions, a given gene is expressed. A screenshot is found in Plate 5.3. All records in PopulusDB has been mapped onto the gene models in the JGI browser (Plate 5.4) making it possible to easily jump between the databases to get both sequence, chromosomal locatization and expression data for about 1/3 of all *Populus* genes.

Expression profiling II – DNA microarrays

It is not cost-efficient to perform EST sequencing to get gene expression patterns in many different samples, but if the ESTs are spotted onto glass slides to make DNA microarrays, many more samples can readily be analysed. Two of the POPYOMICS partners had, prior to the start of the project, created *Populus* DNA microarrays and successfully started to use them to perform transcript profiling in *Populus*. Like in the situation with EST sequencing, large funds and a high number of researchers not participating in POPYOMICS had been involved in this, but we had possibility to use the microarrays in

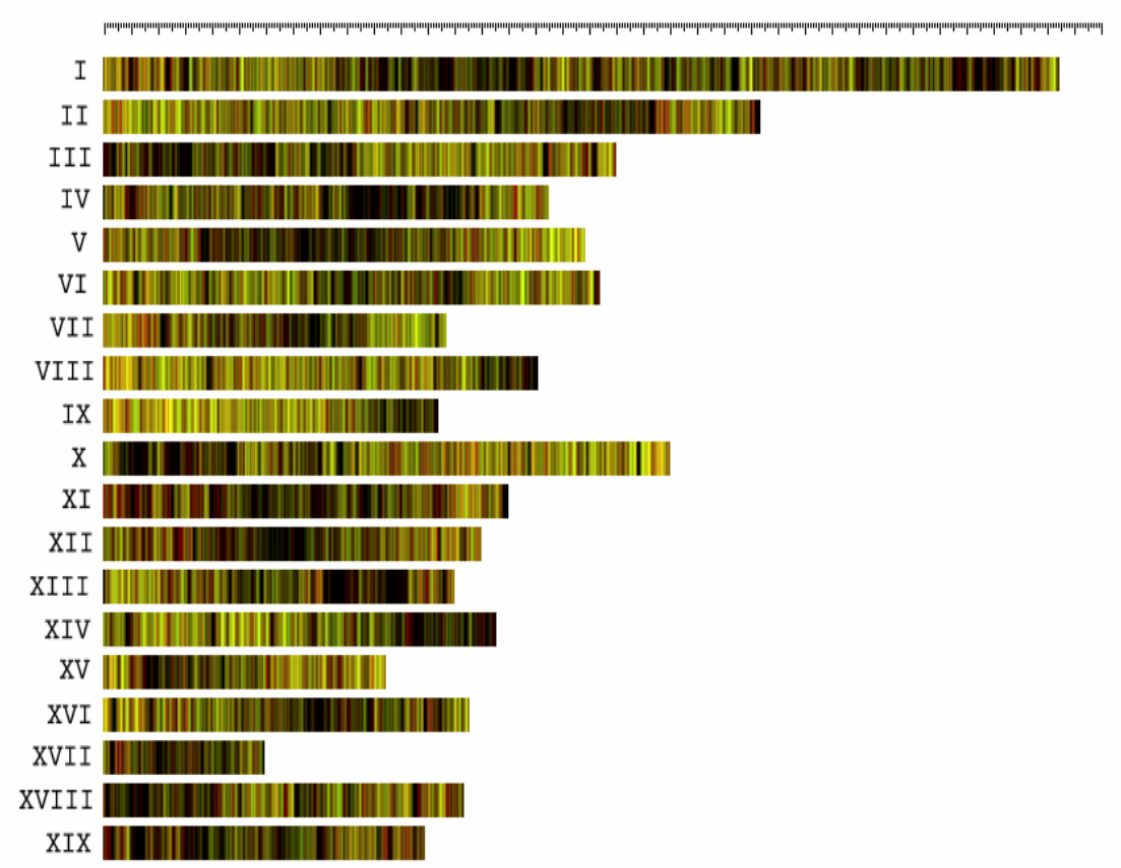


Plate 5.4. Map positions of the 16 500 ESTs on the microarrays used by partner 7. A yellow line shows the map position of gene covered by ESTs, red lines indicate several genes in a 100 000 bp window covered by ESTs.

POPYOMICS subprojects

Significant developments in the methods for DNA microarray analysis were made during the project. Partner 7 used a 25 K microarray whilst partner 8 use a 22 K microarray and for some experiments, other microarrays have also been tested. Most microarrays have been analysed in the lab of partner 7, so this will here be used as an example.The methods for the wet part of microarray analysis, e g RNA extraction, cDNA synthesis and labelleing, manual or automatic hybridization and scanning were all developed and standardized [4]. Likewise, the downstream datamining part, involving e g normalization and quality control, was also much developed and facilitated by the installation of the BASE-software. A very large number of modifications were made to the original BASE software, new functions and plug-ins were created, others were improved, and the resulting database, UPSC-BASE, (www.upscbase.db.umu.se) is not only a data depository but a full-featured DNA microarray analysis tool [5]. This work has in part been funded by POPYOMICS and in part by other sources.

A screenshot of UPSC-BASE is found in Plate 5.5. A lot of standard protocols have been generated, plug-ins have been generated and an analysis pipeline has been established. UPSC-BASE has two domains; all data from published experiments, and some non-published, are found in the public domain. The POPYOMICS, and other unpublished, microarray data, can only be accessed by the persons involved in generating the data.

Genomic technologies – concluding remarks

The major *Populus* genomic tools – genome sequence, EST databases and DNA microarrays - has been generated by using resources from many large sources of funding, but the inclusion of POPYOMICS funding has made it possible for Popyomics researchers to both get early access to these databases but also to get access to restricted areas, therefore making it possible for POPYOMICS to fully exploit the available *Populus* genomic tools with a relative minor input in terms of resources.

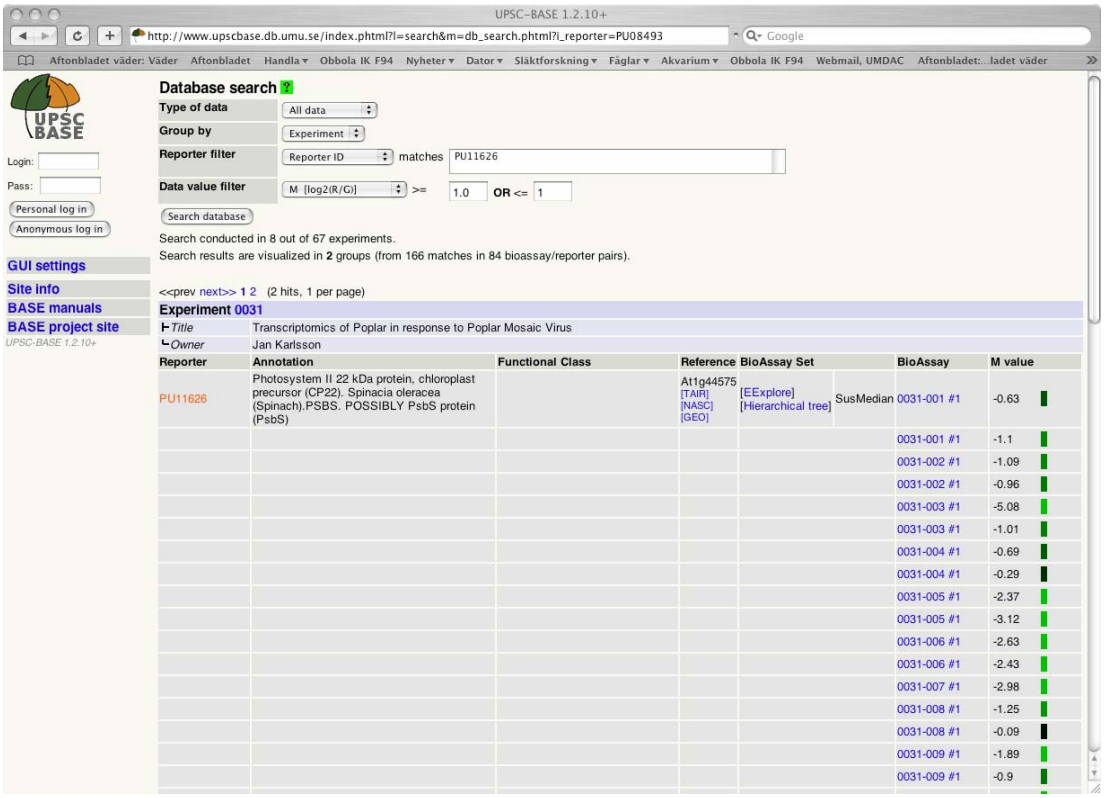


Plate 5.5. Screenshot of UPSC-BASE

5.3 Using Genetical Genomics within POPYOMICS

Genetical genomics to understand Populus adaptation to drought in the POP1 pedigree

Aim

The genetic nature of tree adaptation to drought stress was examined by utilising variation in the drought response of POP1. We combined phenotyping, Quantitative Trait Loci (QTL) analysis, and microarray experiments to demonstrate that "genetical genomics" can be used to provide information on adaptation at the species level. The grandparents, parents, and 167 genotypes of the F_2 population were subjected to soil drying and contrasting responses to drought across genotypes, including for leaf coloration, expansion, and abscission, were observed and QTL for these traits mapped. A subset of extreme genotypes exhibiting extreme sensitivity and insensitivity to drought on

the basis of leaf abscission were defined and microarray experiments conducted on these genotypes. The extreme genotype groups induced a different set of genes; 215 and 125 genes differed in their expression response between groups in control and drought respectively, suggesting species adaptation at the gene expression level. Co-location of differentially expressed genes with drought specific and drought responsive QTLs was examined and these may represent candidate genes contributing to the variation in drought response.

Experimental approach

We used an experimental design that allowed us to identify those genes exhibiting strong segregation associated with drought response within the population (Figure 5.1). Two extreme groups of genotypes (the five highest and lowest) were identified based on the trait 'percentage effect response to drought of the percentage of abscised leaves' and the labelled cDNA obtained from individual replicate leaves from genotypes within each group was pooled for microarray analysis. Our aim was to investigate whether large-scale differences could be identified in the transcriptional response of the two extreme groups. This approach can be considered similar to that of performing a bulk segregant analysis, although with the obvious difference that the data obtained from microarrays is continuous rather than discrete. As the POP2 array had subsequently been developed by P7, containing almost twice as many elements as the POP1 array, we were able to assay differential expression of many more genes in this experiment.

Following on from this array analysis, a group of candidate changes for full 'eQTL' analysis was selected. RNA was extracted from all F_2 progeny and real time Q PCR is being used to investigate the expression of these genes.

Results

Abscission was selected as it is a late onset indicator of drought adaptation response and showed a significant treatment and genotype effect within the F_2 genotypes, with drought-specific QTL being mapped that account for a total of 43% of trait variance.

Analysis of this dataset revealed a striking difference in the transcriptional response of the two extreme groups. When comparing significant changes in response to drought for the high-abscission extreme with those of the low-abscission genotypes, only 65 genes were common to both lists (55 of these being up- and 10 down-regulated). In the high abscission group there were 128 genes with a high probability of differential expression in response to drought, and in the low abscission group 386. This proved that our experimental design using a pooling strategy was capable of detecting segregating patterns of gene expression, allowing us to gain insights into the genomic structure of the transcriptional drought response and showing that a major factor distinguishing the two phenotypic extreme groups was the differential expression of genes between them.

QTL were mapped from the physiological data in control conditions, drought, and for percentage effect response to drought (these being termed 'response' QTL) with percentage effect being the percentage difference between the control and drought condition. In total 25 QTL were mapped in control conditions, 44 in drought, and 30 for percentage effect. Six QTL were common to control and drought (constitutive QTL) and 54 were specific to the drought and/or response to drought. Co-location of drought-specific QTL for a number of traits was observed on Linkage Groups (LGs) IV, VI, VII, VIII, IX, XI, XVIII, and XIX. These all represent regions of the genome that function in controlling drought response with effects on multiple drought response/tolerance related traits, and in the context of this experimental approach may represent likely locations of a regulon-controlling transcription factor. Trait-specific drought/response QTL were also identified for some traits, for example a QTL for height on LG V(I) that explains approximately 8% of the trait variance in both conditions.

Using the sequence of Simple Sequence Repeat (SSR) marker primers as a link between the physical and genetic maps, the positional ratios of genes along a chromosome and of QTL along LGs were calculated as a means to examine co-location (Figure 5.2). This resulted in 13 genes that differed in their expression between extreme groups and that co-located to genomic regions identified by QTL mapping. A number of these genes have known functions that could impact upon the ability to tolerate drought stress or on the

nature of the response initiated upon exposure to drought stress. The poplar genome contains an estimated 40514 genes models. We have calculated that the QTL regions examined (i.e. the entire genome regions within the CI's for all QTL shown) contain 2167 genes and we are examining the co-location of a total of 340 genes. We also know that the POP2 arrays used represent 16,435 gene models. We have therefore calculated that a random list of 340 genes could be expected to contain 7.4 genes that co-locate to our QTL regions. Our results therefore show a greater number of co-locations than would be expected by chance.

All data analysis of transcriptional data was performed in UPSC-BASE (http://www.upscbase.db.umu.se/), where the data will become publically available at the point of publication. UPSC-BASE has been deveopled by P7 and offers a highly efficient and powerful means of analysing transcriptional data obtained using the POP arrays. It also facilitates collaboration through the fact that it is a web-based system.

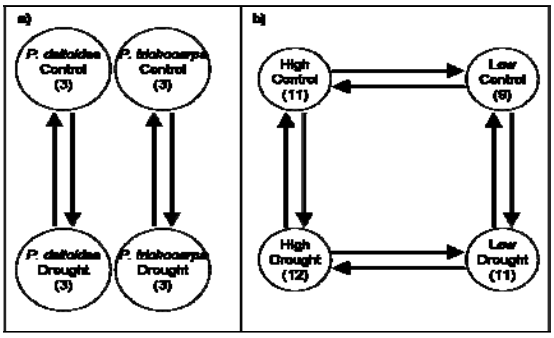


Figure 5.1 Experimental design of microarray experiments. Arrows indicate dye orientations. Replicate arrays conducted with cDNA synthesised from the same RNA extraction but with dyes in the reverse orientation were performed for all comparisons made. **(a)** Microarray experimental design to examine the transcriptional drought response of *Populus deltoides* and *P. trichocarpa* The number of biological replicate arrays is indicated in brackets. **(b)** Microarray experimental design to examine and compare the transcriptional response of extreme F_2 genotype groups for the abscission response in control and drought stress conditions. Each circle represents a pool of cDNA with pools representing the five highest or lowest abscission drought response genotypes sampled in either control or drought conditions.

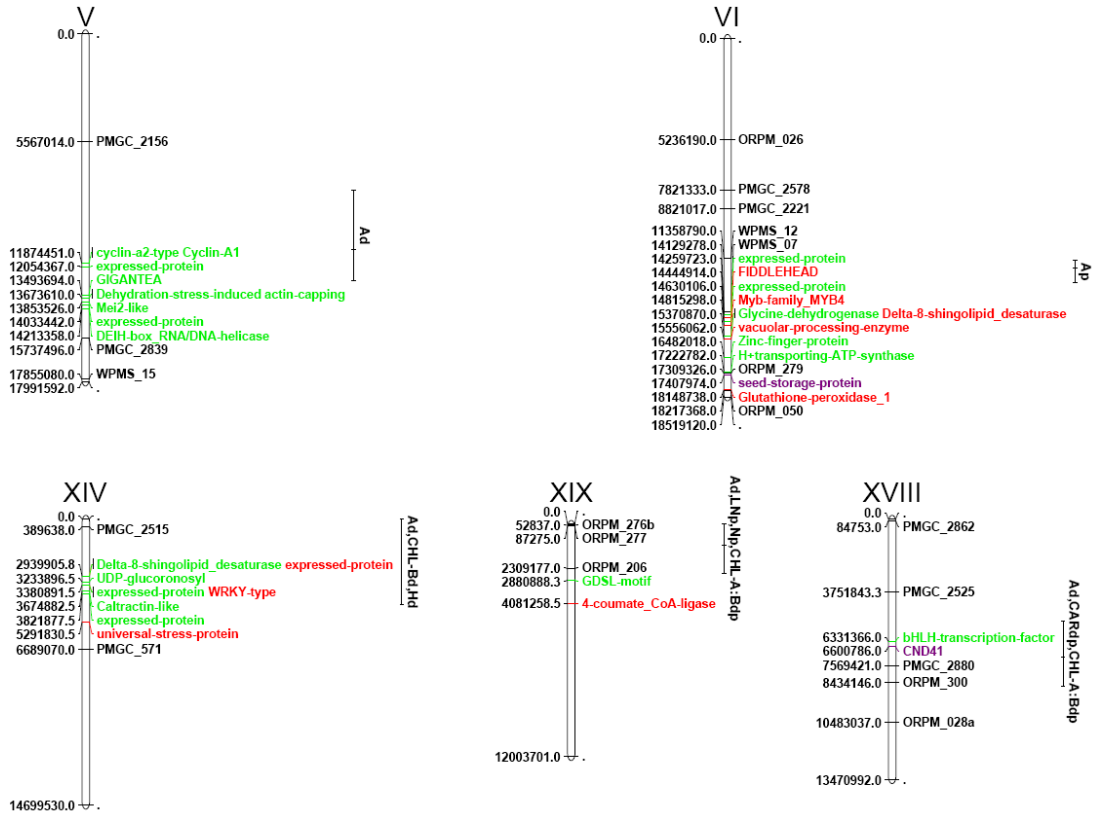


Figure 5.2 The co-location of abscission QTL with genes differentially expressed in response to drought in the grandparental species, *Populus deltoides* and *P. trichocarpa* (red), extreme F_2 genotype groups for the abscission response to drought (green), and both (purple). Each extreme group is composed of the highest or lowest five genotypes for the trait percentage abscission in response to drought. Where QTL for other traits co-locate to abscission QTL, these have been shown. Trait names are in CAPS followed by a lowercase letter indicating treatment conditions. Traits are A – abscission, N – Necrosis, CHL-b – chlorophyll b content (µg ml^{-1}), CHL-a:b – chlorophyll a:b, CAR – carotenoid content (µg ml^{-1}), H – height. Conditions are d – drought, p – percentage effect calculated as ([control-drought]/[control])*100. Only those Linkage Groups (LGs) with abscission QTL identified are shown. Chromosomes are represented by the open bars with the chromosomes number labelled. The values to the left of the bar represent calculated base-pair positions. Gene names and SSR identifiers are shown to the right of the bar. The position of genes and SSRs along a chromosome and of SSRs and QTL along LGs were calculated on a ratio scale and then converted to base-pair positions. Only SSR markers are shown, but QTL were mapped using a genetic map containing additional (non sequence based) markers as detailed in Experimental Procedures section. Confidence Intervals were defined using an F 2 drop off.

Discussion

Rather than simply gaining an overview of the transcriptional response to drought of Populus in general, our aim was to test whether we could identify genes that may contribute to the genetic architectural differences accounting for the contrasting adaptive drought response mechanisms of *P. deltoides* and *P. trichocarpa*. Such an analysis requires stringent statistical treatment and to this end we undertook an ANOVA analysis to determine genes differentially expressed between the two species based on ratio-of-expression values. 569 genes that were differentially expressed between the species were identified, with 59 of these remaining significant after applying a Bonferroni correction. ANOVA was the most suitable statistical test in this case, where no direct connection between species existed on arrays and so ratio-of-expression comparisons were being tested. This list identified genes in regulons that were differently regulated in the two species - regulons that may be important for the different drought responses in the two species. Among the genes more highly expressed during drought in *P. trichocarpa* were those encoding blight-associated protein P12, a putative disease resistance protein, a ripening-related protein, an expansin (with an apparent leaf-specific expression pattern, deduced from the library distribution of EST clones), a methyl transferase, and an epimerase/dehydratase. Genes with an informative annotation that were more highly induced in *P. deltoides* included those encoding enzymes in starch metabolism, a granule-bound glycogen/starch synthase, a 1-4-alpha-glucan branching enzyme (starch branching enzyme), and proteases (ClpR1 and ubiquitin). In addition, a 1-aminocyclopropane-1-carboxylate oxidase (ACC oxidase) was induced in *P. deltoides* to a considerably greater extent than in *P. trichocarpa*. There are many ACC oxidase genes expressed in *Populus* and data-mining of PopulusDB showed that these genes often had drastically different expression patterns; this particular ACC oxidase gene (PU10422, PU10214 Table II), appears to be almost exclusively expressed in senescing leaves. Ethylene plays a key role in controlling the senescence and abscission of leaves, both in response to stress and in autumnal senescence, and it is therefore interesting that in these two species ethylene was more highly induced in the species exhibiting senescence in response to drought.

Although the commonality in the transcriptional responses of the two species to drought was large, the data presented here indicate that there were a number of genes that differed in expression in response to drought between the two *Populus* species, suggesting that at least a part

of their contrasting physiological responses to drought may be genetically controlled through differences in gene regulation. To provide further insight into this, we performed a drought stress experiment using the F_2 offspring of the grandparent clones.

Transcriptional separation of drought response within the population. In order to assess the extent to which gene expression may account for the separation of F_2 genotypes in their drought response, ideally the transcriptome of all individuals in the population should be analysed during drought stress, as was undertaken for a population subset [1] to examine eQTL in wood-forming tissues. Rather than assaying our entire population, we attempted to construct an experimental design that allowed us to identify those genes exhibiting strong segregation associated with drought response within the population. Two extreme groups of genotypes (the five highest and lowest) were identified based on the trait 'percentage effect response to drought of the percentage of abscised leaves'. Our aim was to investigate whether large-scale differences could be identified in the transcriptional response of the two extreme groups, and we felt that a pooled loop design provided the most efficient strategy for testing this. Although we are aware of the limitation of such a pooling strategy (we have no indication of the contribution of individual replicates or genotypes within the pool), we do not feel that this represents a significant limitation to our approach: our aim is to identify those genes with the most highly segregating expression levels between the two sets of extreme genotypes. The pooling strategy will achieve this goal but may also identify a proportion of false-positives that achieved significance through the influence of an individual replicate or genotype on the mean expression represented by the pooled sample. Such a degree of false positives is acceptable as they can readily be screened through subsequent Real-Time Polymerase Chain Reaction (RT-PCR) confirmation before progressing on to eQTL mapping within the entire population. We are currently undertaking this confirmation and subsequent eQTL mapping strategy. This approach can be considered similar to that of performing a bulk segregant analysis, although with the obvious difference that the data obtained from microarrays is continuous rather than categorical.

Abscission was selected as it is a late onset indicator of drought adaptation response and showed a significant treatment and genotype effect within the F_2 genotypes, with drought-specific QTL

being mapped that account for a total of 43% of trait variance. Although percentage abscission was the most significant separator of the extreme groups, relative leaf area expansion rate and carotenoid : total chlorophyll content also showed significant differences.

Analysis of this dataset revealed a striking difference in the transcriptional response of the two extreme groups. Again, differentially expressed genes were identified using the ebayes method of the 'LIMMA' package in R. When comparing significant changes in response to drought for the high-abscission extreme with those of the low-abscission genotypes, only 65 genes were common to both lists (55 of these being up- and 10 down-regulated). In the high abscission group there were 128 genes with a high probability of differential expression in response to drought, and in the low abscission group 386. This proved that our experimental design using a pooling strategy was capable of detecting segregating patterns of gene expression, allowing us to gain insights into the genomic structure of the transcriptional drought response and showing that a major factor distinguishing the two phenotypic extreme groups was the differential expression of genes between them.

The loop design employed allowed data to be interrogated from a high vs low extreme group perspective in both control and drought conditions. In drought there were 48 genes more highly expressed in the high abscission extreme group and 77 genes with lower-expression. In this comparison, a gene with a higher expression level in the high abscission extreme group than in the low will have a positive $\log_2$ ratio value, and a gene with higher expression in the low abscission extreme group will have a negative $\log_2$ ratio. In control conditions, 96 genes were more highly expressed and 119 showed lower expression. Such control-condition expression differences represent constitutive differences in gene expression across the population that could contribute to drought susceptibility and tolerance.

The functional role of genes based on GO categories (as described above) was examined. Table III gives GO categories that were significantly over-represented by differentially expressed genes in either the high abscission or low abscission extremes in control and drought conditions. These data allow two questions to be asked: 1) Are there classes of genes that differ in their constitutive

expression and could account for differences in drought response? 2) Are there drought responsive gene classes that differ in their degree or direction of expression change that could account for the differential drought response? In control conditions, high abscission extremes had higher expression levels of genes involved in hormonal response and signalling, including ABA mediated signalling and response, jasmonic acid and ethylene mediated signalling, and categories involved in biotic stress and pathogenesis responses. The expression differences for ABA signalling, biotic stress, and pathogenesis responses remained significant in drought conditions. As for the grandparental gene expression responses, it is interesting that ethylene signalling was more highly expressed in the high abscission extremes. It is also of note that a number of biotic stress and pathogenesis response categories are significantly more highly expressed in high abscission lines, as this may suggest that these genotypes initiate a response that has a greater overlap with the Hypersensitive Response-type response of plants. Indeed, these genotypes regulate genes involved in Reactive Oxygen Species (ROS) removal such as glutathione peroxidase and superoxide dismutase (SOD) to a lesser extent than low abscission response genotypes.

Even more striking differences were revealed when examining genes that were more highly expressed in the low abscission extremes. In control conditions many categories involved in secondary metabolite synthesis including phenylpropanoid biosynthesis, flavonoid biosynthesis, anthocyanin biosynthesis, chalcone biosynthesis, and monoterpenoid biosynthesis showed significantly higher expression in the low abscission extremes. There were also categories involved in cell wall modification, organisation and pattern definition. The apparent emphasis on homeostasis and cellular protection may indicate that resources are being utilised to maintain the cellular integrity and biochemical functionality of leaves. For example, many of these categories could result in an enhanced ability to tolerate increases in reactive oxygen species and other drought-induced biochemical stresses.

In drought conditions many categories involved in cellular homeostasis maintenance were represented, as were categories for gibberellic acid (GA) response and signalling, suggesting a

functional role for this hormone in the drought response of trees that do not readily abscise leaves when exposed to drought stress.

We then investigated the library distribution of the regulons that were differentially expressed, either in drought or in control conditions, in drought-sensitive vs. drought-resistant clones in the population. As for the common responses to drought in the grandparents, we studied the digital expression profiles in PopulusDB for the gene lists identified above. Within the high abscission extremes there was no strong pattern of library distribution, perhaps as a result of the small number of differentially expressed genes (data not shown). More obvious patterns were seen for the low abscission extremes, with the greatest number of up-regulated genes being located within the dormant bud and dormant cambium tissue libraries perhaps reflecting the adaptive response of these genotypes. A combined list of genes differing in their expression between the high and low extremes in drought and control conditions was gathered and analysed for selection of candidate genes for eQTL assessments.

Integrating transcriptional and QTL data. We wanted to examine the degree to which differentially expressed genes co-locate to genomic regions identified by QTL analysis. If co-location occurs, this could be due to differences in *cis*-acting elements (promoter sequences) in the differentially expressed genes. This increases the probability that a gene is involved in the control of drought response, although it does not prove a causal link since the expression may be regulated in *trans* by changes in e.g. a transcription factor regulating the drought response.

Both the grandparental and extreme genotype datasets provide candidates for explaining segregation of drought response at the transcriptional level, although not all grandparental differences would segregate within the population. As discussed above, there were considerable differences between the two sets of candidate genes, with a greater number of co-locating genes being identified from the extreme genotype comparison.

Thirteen genes that differed in their expression between extreme groups and that co-located to genomic regions identified by QTL mapping were identified. A number of these genes have known functions that could impact upon the ability to tolerate drought stress or on the nature of the response initiated upon exposure to drought stress.

We are now undertaking work to map the expression of these genes, and other genes identified from this study, within the F_2 population and believe that the approach undertaken here has provided a highly efficient means through which to identify the genes of greatest interest and potential for eQTL mapping.

Conclusions

We have shown that *P. deltoides* and *P. trichocarpa* have contrasting responses to drought, and have used a genetic and genomic approach to study the genetics of this difference. We have shown that the divergent drought response of two *Populus* species exhibits transgressive segregation within an F_2 population and that this results in the emergence of highly contrasting adaptive drought responses. We observed dramatic segregation within the population in the abscission response to drought and the comparison of the transcriptional response of a set of high and low abscission genotypes revealed a striking and perhaps surprising degree of separation. Importantly, we have also shown that an experimental design using few microarrays serves as an efficient means to identify genes with segregating patterns of gene expression within a population.

Although the common transcriptional responses to drought stress provides information on genes and regulons induced by drought stress (i.e. genes typically expressed by dormant tissues), this type of study does not necessarily provide good candidates for genes responsible for natural variation in this trait, since it is quite likely that the different alleles of the genes regulating the drought response may not have particularly contrasting transcript levels. Instead, we believe that the "genetical genomics" approach may be needed to understand the drought response at the level of genetic variation. A number of genes with differential levels of expression response to drought between the two extreme groups co-located to genomic regions identified by QTL

analysis. These genes may provide clues as to the genetic mechanisms through which species adaptation to drought has been achieved. A similar study in rice identified large-scale divergence of the transcriptional response for genotypes with divergent osmotic adjustment responses to drought stress (Hazen *et al.*, 2004) and this may suggest that gene expression control plays a key role in the mechanisms of species divergence.

Genetical genomics in ozone

Aim

The depletion of stratospheric ozone has been well documented, but during the last century tropospheric ozone concentrations have risen from 10-15 ppb to 30-40 ppb [6] Furthermore, [8] predicted that up to 50% of forests will be exposed to 60 ppb by 2100; a concentration of this pollutant that is known to damage wild and cultivated plants alike, and has had profound impact on a number of forest ecosystems, first noted in the San Bernardino mountains [9] More recently, [10] conducted a long term study of the effects of O_3 pollution on several tree species, and demonstrated a negative impact throughout the whole ecosystem, not only on the trees themselves but also on pests and soil micro-organisms. The highly destructive and costly effects of ozone highlight the importance of gaining an understanding of the genetic basis of the mechanisms of ozone damage in plants, so that adaptation may be understood and tolerance improved through selection and breeding strategies such as marker-assisted selection, or through genetic modification.

Experimental approach

The grandparents, parents, and167 genotypes of the F_2 population were subjected to enriched ozone levels using Open Top Chamber (OTC) facilities and contrasting responses to ozone across genotypes, including for leaf coloration, expansion, and abscission, were observed and QTL for these traits mapped.

Plate 5.6 Open top chambers at Headley, and the range of damage severity between genotypes.

167 genotypes of POP1 were grown in Open Top Chambers (OTCs) and exposed to either ambient or enriched ozone (100 ppb) for three months. For all traits, there was a highly significant effect of genotype (p<0.0001). No significant ozone effect on height was found. For basal stem diameter a small (-3.2%) but significant negative effect ($F_{1,132}$=11.24, p<0.001) of ozone treatment was found. Ozone had a positive influence on the total number of leaves initiated both early ($F_{1,123}$=4.62, p<0.05) and late ($F_{1,126}$=5.45, p<0.05) in the season. The area of the first unfurled leaf was significantly greater in ozone ($F_{1,93}$=6.82, p<0.01), whilst leaf area expansion rate was significantly reduced ($F_{1,93}$=7.82, p<0.01). No significant effect on final leaf area was found, although the genotype*treatment interaction was almost significant ($F_{93,274}$=1.25, p=0.09). There was a highly significant increase in leaf abscission. Chlorophyll content showed no significant treatment effect, but a significant genotype*treatment interaction ($F_{126,416}$=1.27, p<0.05), with some genotypes showing an increase in chlorophyll content in response to ozone, and others showing a decrease. These results are documented in Table 5.1.

A group of three sensitive and three tolerant genotypes were selected on the basis of visible damage and necrotic lesion development after both 30 and 70 days of treatment with 100 ppb ozone in open-top chambers. The trait means and the outcomes of statistical tests are represented in Table 5.2. No significant differences are seen between sensitivity group for any of the traits except those representative of visible damage at 30 days and 70 days, and those relating to leaf

size. The selected genotypes were exposed to a 9 hr treatment of 200 ppb ozone in growth chambers, and the photographic record of symptom progression is presented in figure. The sensitive genotypes showed more severe necrotic damage than *P. trichocarpa*, whilst the tolerant clones remained symptom free (Plate 5.6). We had previously exposed the grandparental species to enriched ozone using growth chamber facilities. Plants were exposed to enriched ozone (200 ppb) for 9 hourrs at which time leaf material was sampled for RNA extraction and microarray analysis.

Results

This study identified the 50 genes most differentially expressed between species, using PLS-DA. A notable species specific expression pattern is for *xyloglucan endotransglycosylase (XET)*, with three ESTs that are strongly down-regulated in *P. deltoides*, but only moderately so in *P. trichocarpa*. XETs have been shown to cleave cross-linking glycans in cell walls, and are involved in cell expansion [11]. A species difference in XET protein levels in response to ozone could be indicative of a differing response of cell expansion and therefore leaf growth.

Of the 50 most ozone responsive genes, 11 belong to the protein synthesis GO category. These mainly consist of ribosomal proteins and are generally up-regulated, particularly in the sensitive group. This may be indicative of an increase in overall protein synthesis in response to ozone stress, and could be due to an increased demand for stress-related proteins, and also in order to counter ozone induced protein degradation. Changes in expression of stress-responsive categories (response to abiotic stimulus, biotic stimulus and oxidative stress) are expected, and are consistent with literature describing similarity between the response to ozone and to pathogens. When compared to the results for *P. trichocarpa* and *P. deltoides*, there are some similarities, with representation of genes from stress-related categories appearing in both lists. ESTs encoding peroxidase and kunitz trypsin inhibitor appear in both lists, lending support to their involvement in the response to ozone.

Five replicates of five sensitive and five tolerant clones were defined from the OTC experiment and later grown and exposed to either ambient or ozone-enriched air (200 ppb) using growth chambers. RNA was extracted from sampled leaf material after 9 hours of exposure to enriched

ozone. Equal amounts of RNA from the five replicates of each clone for each treatment were pooled to give 100 μg RNA per treatment per clone to be used for cDNA synthesis and microarrays were then performed to examine the response to ozone for each genotype. Lesion morphology was similar in both the acute growth chamber study and chronic OTC exposures, indicating a similar mode of symptom development. The separation between the groups was considerably greater than that between *P. trichocarpa* and *P. deltoides,* with lesion formation being more rapid and extensive in the sensitive F_2 genotypes, indicating that the population also exhibits transgressive segregation for the response to acute ozone treatment. The investigation therefore successfully identified two groups of genotypes with greater separation than the grandparents for ozone-induced visible damage, and as such may be useful as a model system for identifying the mechanisms of lesion formation. PLS-DA analysis of the microarray data successfully identified a group of 50 genes that were differentially expressed between sensitivity groups (Figure 5.3). Examination of the expression patterns of these genes reveals that there are larger changes in gene expression in the sensitive group. This may be interpreted as the genotypes in the tolerant group not perceiving the ozone treatment as a stress, whilst the sensitive group underwent transcriptional changes associated with a defence response.

QTL were mapped from the physiological data collected during the OTC chronic ozone exposure. Of the 58 QTL detected, 20 were found to map only under ozone treatment or as response QTL. It is therefore likely that these represent genomic regions that are specifically involved in governing the response to ozone, and are the most encouraging candidates for further investigation. Such QTL were found for leaf necrosis, diameter, late-season leaf number, height, late season abscission, area of the first unfurled leaf and chlorophyll content, indicating that ozone responsive genomic regions exist that govern numerous traits. QTL that map both in ozone and control conditions are likely to correspond to regions that are of more fundamental importance in governing the trait under differing environmental conditions.

a)

	P. deltoides			*P.trichocarpa*		
	Control	**Ozone**	**% effect**	**Control**	**Ozone**	**% effect**
Height (cm)	55.30	53.00	-4.34	79.30	72.30	-9.68
Diameter (mm)	6.03	5.43	-10.00	6.47	5.17	-20.20
Leaf number (30d)	15.70	14.70	-6.80	19.70	15.30	-28.76
Leaf number (70d)	21.30	18.30	-16.39	26.70	26.00	-2.69
Chlorophyll content (CCM reading)	15.10	9.53	-58.45	20.70	21.90	5.48
% abscised (30d)	7.10	17.60	59.66	8.80	10.30	14.56
% abscised (70d)	18.50	40.00	53.75	22.80	53.90	57.70
Area of first unfurled leaf (cm^2)	7.40	6.64	-11.38	9.94	12.62	21.20
Leaf expansion rate (% area increase)	74.60	68.20	-9.38	87.20	54.40	-60.29

b)

	P. deltoides			*P.trichocarpa*		
	Control	**Ozone**	**% effect**	**Control**	**Ozone**	**% effect**
Height (cm)	55.30	53.00	-4.34	79.30	72.30	-9.68
Diameter (mm)	6.03	5.43	-10.00	6.47	5.17	-20.20
Leaf number (30d)	15.70	14.70	-6.80	19.70	15.30	-28.76
Leaf number (70d)	21.30	18.30	-16.39	26.70	26.00	-2.69
Chlorophyll content (CCM reading)	15.10	9.53	-58.45	20.70	21.90	5.48
% abscised (30d)	7.10	17.60	59.66	8.80	10.30	14.56
% abscised (70d)	18.50	40.00	53.75	22.80	53.90	57.70
Area of first unfurled leaf (cm^2)	7.40	6.64	-11.38	9.94	12.62	21.20
Leaf expansion rate (% area increase)	74.60	68.20	-9.38	87.20	54.40	-60.29

Table 5.1 Summary of traits measured in **a)** *P. deltoides* and *P. trichocarpa*; and **b)** the F_2 population in response to a chronic 100 ppb ozone treatment. Where traits were measured on more than one occasion, this is indicated by 30d and 70d (30 or 70 days of exposure). The data represent means for each trait recorded in ambient conditions and elevated ozone, and the percentage effect of ozone upon the trait. A general linear model was used to analyse all data except *%abscised (30d)* where a Kruskal-Wallis test was performed due to a non-normal distribution. Significance levels are indicated by * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Discussion

An aim of this study was to identify candidate genes through co-location of differentially expressed genes from microarray analysis with QTL identified from the population experiment. To achieve this, the top 50 most differentially expressed genes between *P. deltoides* and *P.*

trichocarpa, and also between the groups of sensitive and tolerant F_2 genotypes, were located on the physical sequence. The physical and genetic maps were then aligned using sequenced SSR markers.

An EST with high similarity to Arabidopsis *arginine decarboxylase 2* (*ADC2*) was found to be significantly differentially expressed between *P. deltoides* and *P. trichocarpa*, and other ESTs representing this gene were also found to exhibit a similar pattern. Interrogation of the *P. trichocarpa* genome revealed this gene to be located on linkage group IV, and is coincident with a QTL for late-season visible damage under ozone treatment. As discussed, co-location of a differentially expressed gene with a QTL region provides additional evidence for the gene being important in governing the response [17]. The maternal and paternal effects for the QTL were both negative (data not shown), indicating that the QTL alleles from both parents serve a protective role against ozone damage. If the gene underlying the QTL is indeed *ADC*, this observation is consistent with the literature which has shown polyamine synthesis, and arginine decarboxylase itself, to ameliorate ozone damage [18, 19]. In particular, application of a specific inhibitor of ADC resulted in increased visible symptom formation in ozone exposed barley [19]. A role for polyamines and ADC has been postulated for tolerance to a number of other stresses, including salt and osmotic stress [20] and chilling [21]. It is of note that *ADC* is more up-regulated in *P. trichocarpa*, the ozone sensitive species. This could be considered as evidence against a key role for this gene in determining ozone sensitivity. However, it must be emphasised that alleles with a protective effect may be present in the sensitive parent and vice versa, and that this is considered a likely mechanism for transgressive segregation in the F_2 population.

	Control		Ozone							
	Tolerant	Sensitive	Tolerant	Sensitive		G	T	S	T*S	T*G
% damaged 30d	-	-	1.36	19.54		-	-	*	-	-
% abscised 30d	0.53	2.89	11.26	14.08		ns	*	ns	ns	**
% damaged 70d	-	-	24.05	40		-	-	**	-	-
% necrotic	-	-	0.18	23.23		-	-	**	-	-
% abscised 70d	11.84	10.11	16.08	29.25		ns	ns	ns	ns	**
Height	88.33	78.58	77.33	88.28		ns	ns	ns	ns	ns
Diameter	8	6.94	6.53	6.97		ns	ns	ns	ns	ns
Chlorophyll content	20.34	20.58	24.37	33.34		*	*	ns	ns	ns
Leaf area 1	9.52	6.4	7.91	12.34		**	**	ns	***	ns
Leaf area 2	22.21	13.31	14.14	21.8		*	ns	ns	**	ns
% area increase	134	116	81	84		ns	*	ns	ns	ns
Leaf number 30d	23.45	21	20.5	23.87		ns	ns	ns	ns	ns
Leaf number 70d	30.81	28	29.41	32.2		ns	ns	ns	ns	ns

Table 5.2 Means of physiological traits for the selected sensitive and tolerant clones of Family 331, identified from a 100 ppb ozone exposure in open top chambers. For traits measured on more than one occasion, this is indicated by 30d (30 days of exposure) and 70d (70 days of exposure). Results of ANOVA are given, with * $p<0.05$, ** $p<0.01$, *** $p<0.001$. G= genotype, T= treatment, S=sensitivity group, T*S= interaction between treatment and sensitivity group, T*G= interaction between treatment and genotype

S-adenosyl methionine (SAM) is the precursor molecule to both polyamines and ethylene, both of which are thought to be involved in the response to ozone. It is of interest, therefore, that *SAM synthetase* (*SAMS)* is co-located to QTL for both late and early season visible damage on LG II [14] put forward a hypothesis for the interacting role of polyamines and ethylene in the control of senescence, with ethylene acting as a positive regulator and polyamines as a negative regulator. The authors put forward a model in which polyamine and ethylene biosynthetic pathways compete for a limited pool of SAM, with the interaction between the two determining the outcome. They also postulated that the product of one pathway could act to inhibit the opposing pathway. This presents the intriguing possibility that a similar mechanism could exist in the response to ozone, with the interaction between polyamine and ethylene biosynthesis serving to determine the extent of visible damage. The greater up-regulation of enzymes in the ethylene biosynthetic pathway for *P. trichocarpa* may be indicative of increased ethylene synthesis in this species. If this is the case, the second hypothesis of [14] could be of relevance, as increased ethylene biosynthesis may serve to negatively regulate polyamine biosynthesis. If *SAMS* is indeed the gene underlying the QTL, it is possible that it plays an important role in this interaction through governing substrate availability.

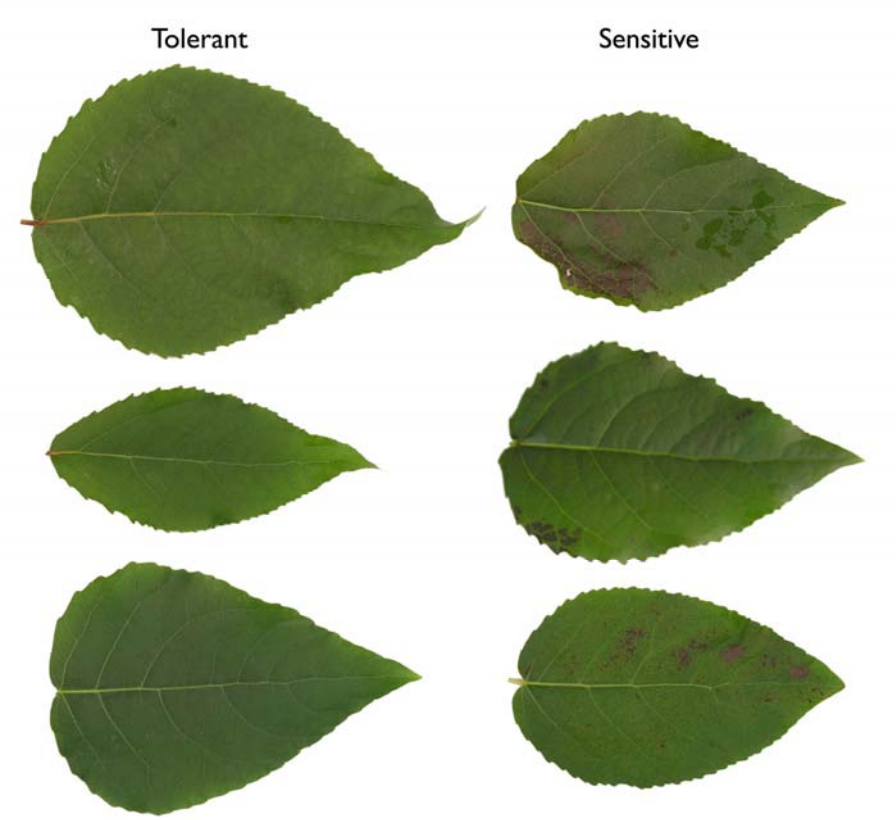


Plate 5.7 The response of three ozone tolerant and three sensitive F_2 genotypes of family 331 to a 9 h exposure of 200 ppb ozone in growth chambers, showing necrotic lesions in the sensitive clones.

Plasma membrane intrinsic proteins are a subset of aquaporins that are located in the plasma membrane. Aquaporins are involved in water transport, and have achieved considerable attention concerning their role in response to abiotic stresses, in particular drought. Recently, it has been found that they are capable of transporting the signal molecule H_2O_2 across the plasma membrane, and as such may represent an interface between ROS generation in the apoplast and intracellular signalling [15]. This opens up the possibility that they are involved in the response to ozone through transportation of H_2O_2 resulting from ozone exposure.

A wound inducible chitinase homologue co-locates to the same visible damage QTL as *ADC*. Chitinases are thought to play a role in defence against invading pathogens, and have also been found to be inducible by wounding in poplar [16]. As chitinase specifically acts to degrade chitin present in fungal cell walls, it is hard to envisage how it would play a role in the response to ozone. As mentioned, the QTL regions identified are large, meaning that co-location to differentially expressed genes can occur by chance, by virtue of being linked to a gene or genes that do play a causal role. Therefore, although co-location of differentially expressed genes with QTL provides evidence for their role, it does not prove a definite causal link. This emphasises the importance of an integration of these approaches with biological understanding. The value of such a multifaceted approach has been described in [17].

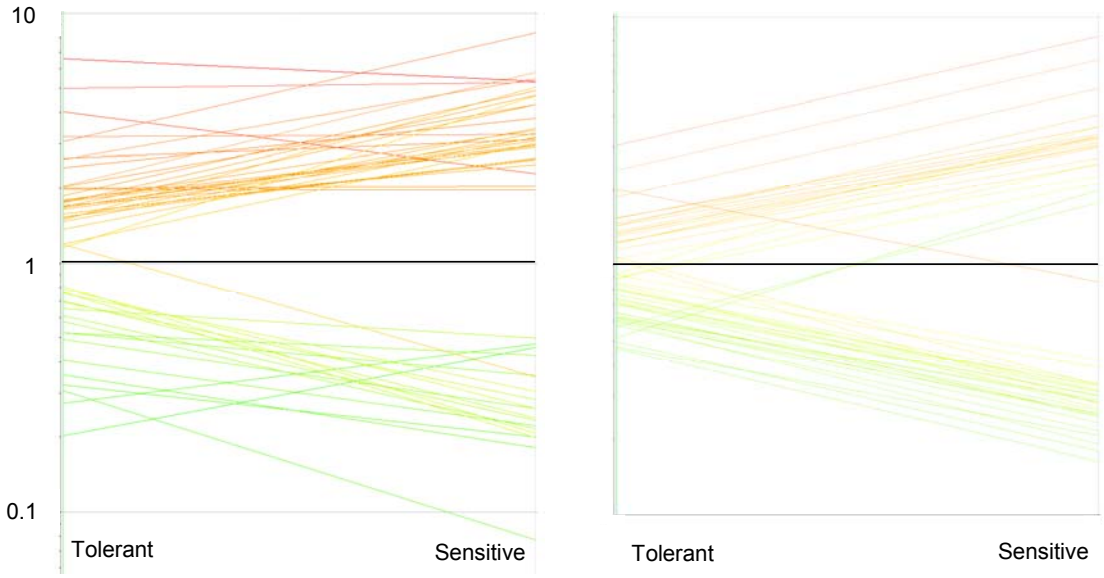


Figure 5.3 Gene expression patterns in groups of three sensitive and three tolerant genotypes of POP1 exposed to 200 ppb ozone for 9 hours in growth chambers. Red colouration indicates up-regulated in elevated ozone, yellow no change, and green down-regulated. The mean expression ratios for the tolerant genotypes are shown on the left of each figure and sensitive on the right, with a line joining corresponding genes. a) The 50 genes most influenced by ozone treatment irrespective of sensitivity group; b) The 50 genes showing the greatest differential expression between sensitivity groups. All analysis was performed using PLS-DA on normalised data.

Genetical genomics on rust in the POP3a pedigree.

Aim

Melampsora larici-populina is the most damaging pathogen for poplar culture in Europe. In the previous section on QTL analysis of quantitative disease resistance, QTLs for various parameters related to quantitative resistance against *M. larici populina* were mapped. A strong QTL on LG XVI of *P. nigra* explained 24% of the variation in diameter of the uredinia of *M. larici-populina* race E4 in pedigree POP3a. In order to identify putative genes responsible for this variation in size, a genetical genomics approach was chosen.

Experimental approach

In our genetical genomics approach, we aim at identifying a gene that 1/ maps in the QTL for uredinia size, 2/ that is differentially expressed between genotypes with small uredinia size and genotypes with large uredinia size and 3/ of which the differential expression is regulated *in cis*. To achieve this, we have grouped the clones of POP3a in two classes, the first class having the QTL allele for large uredinia size and the second class having the alternative *P. nigra* allele. Subsequently, the 10 most extreme individuals of the class with the QTL allele for large uredinia size were chosen. Similarly, from the class with the alternative allele, only those 10 were chosen that had the smallest uredinia size. Subsequently, 6 ramets of these 10 clones were grown in the greenhouse, and 2 leaf disks of each plant were frozen from leaf 7 before infection, from leaf 8 after mock infection with water and from leaf 9 after infection with *M. larici-populina* race E4. RNA was prepared from both sets of leaf samples and used in microarray analyses. Microarray analyses were carried out with the POP3 array.

Results

RNA was extracted from the 2 x 10 samples differing in uredinia size, and this was done for the samples that were not infected and the samples that were infected. The 10 genotypes with largest uredinia size were labeled L1, those with the smallest uredinia size were labeled L2.

L1 = large surface	
K1	95001/42
K2	95001/29
K3	87001/187
K4	95001/127
K5	95001/49
K6	87001/272
K7	97001/46
K8	95001/88
K9	95001/20
KX	97001/49

L2 = small surface	
K1	87001/277
K2	95001/4
K3	87001/17
K4	87001/49
K5	87001/48
K6	87001/41
K7	87001/275
K8	95001/58
K9	87001/260
KX	87001/166

Table 5.3 Two extreme groups were identified, with lagre and small uridinia

The RNA was arrayed in two ways. Firstly, an individual RNA loop was chosen for the samples that were not infected, according to the following design:

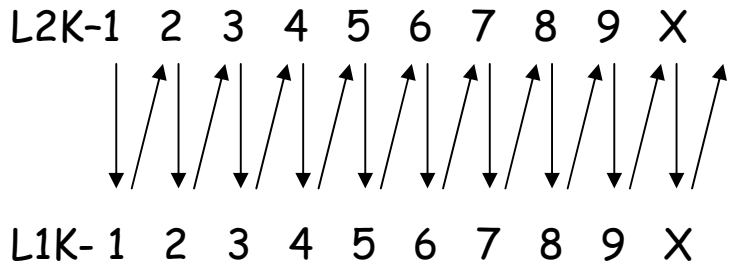


Figure 5.4 Individual hybridization scheme between L1 and L2.

Secondly, the individual RNA samples were pooled to 4 different super-pools containing the same amount of RNA from each individual sample. The super-pools were L1s, L1i, L2s and L2i. These arrays were hybridized according to the following hybridization design.

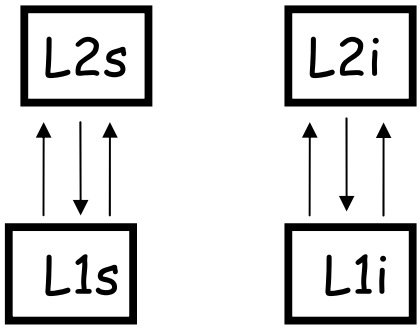


Figure 5.5 Pools hybridized from RNA created from ten individual extractions of L1 (1-10) and L2 (1-10) RNA. (s), before infection and (i,) after infection.

The array data were imported into UPCS-BASE for data storage and statistical analysis. Complete array datasets for all arrays will be available online as experiment UMA-0052 at (http://www.upscbase.db.umu.se). Subsequently, we examined which genes were highest or lowest expressed in the different samples. The top 20 results from the B-Statistics for the experiment when the L1K and L2K arrays were treated as only two groups are given below.

PU.name	Annotation - Up-regulated in small size non-infected samples	M	A	t	P.Value	B
PU09119	Glucan endo-1,3-beta-glucosidase, basic vacuolar isoform [Precursor] PC	3.44	8.67	12.30	3.98E-08	16.90
PU20289	Thaumatin. Vitis riparia (Frost grape) (Vitis vulpina). POSSIBLY osmotin p	3.42	8.34	8.14	2.04E-06	9.22
PU10330	Beta-1,3-glucanase. Hevea brasiliensis (Para rubber tree).HGN1. POSSIE	3.24	7.4	7.52	5.22E-06	7.91
PU29527	expressed protein	3.09	7.24	9.65	2.20E-07	12.20
PU09894	Ypr10 protein. Castanea sativa (Sweet chestnut).YPR10. POSSIBLY polle	3.08	7.44	14.70	3.34E-09	20.20
PU31100	Salivary proline-rich protein 1 (Fragment). Homo sapiens (Human). POSS	2.98	9.54	7.35	6.86E-06	7.46
PU20185	Stress and pathogenesis-related protein. Fagus sylvatica (Beechnut).PR1	2.82	9.84	9.44	3.02E-07	11.90
PU10650	R-LORF10. Turkey herpesvirus.R-LORF10.	2.78	8.95	9.92	1.61E-07	12.80
PU30057	Disease resistance protein. Lycopersicon esculentum (Tomato).CF-5. PO	2.71	8.13	9.28	3.85E-07	11.60
PU26670	cinnamoyl-CoA reductase	2.62	7.05	11.80	4.33E-08	15.80
PU11518	expressed protein	2.62	7.15	11.60	4.33E-08	15.70
PU03804	Class I chitinase. Medicago sativa (Alfalfa).CHIA1. POSSIBLY basic chitin	2.56	9.27	12.90	2.01E-08	17.90
PU10987	disease resistance protein, putative	2.56	7.72	10.50	8.27E-08	13.80
PU29234	No annotation avaliable	2.52	7.78	8.36	1.49E-06	9.70
PU29350	Glycosyltransferase (EC 2.4.1.-). Methanosarcina mazei (Methanosarcina	2.52	8.43	7.78	3.45E-06	8.45
PU30470	Homoglutathione synthetase. Phaseolus vulgaris (Kidney bean) (French b	2.49	7.85	11.30	4.33E-08	15.20
PU11342	Glutathione S-transferase GST 18 (EC 2.5.1.18) (Fragment). Glycine max	2.46	9.93	9.54	2.56E-07	12.10
PU21589	expressed protein	2.45	8.18	7.38	6.58E-06	7.58
PU09965	Blight-associated protein p12 precursor. Citrus jambhiri (rough lemon). PC	2.43	7.33	8.59	1.06E-06	10.20
PU25184	expressed protein	2.40	7.53	8.22	1.81E-06	9.41

a) Before infection up-regulated.

PU.name	Annotation - Down-regulated in small size non-infected samples	M	A	t	P.Value	B
PU12055	ADR11 protein (Fragment). Glycine max (Soybean).ADR11.	-2.74	7.12	-9.69	2.06E-07	12.30
PU08977	Auxin-binding protein ABP20 precursor. Prunus persica (Peach).ABP20. F	-2.22	7.87	-8.48	1.25E-06	9.96
PU22186	LSDV142 putative secreted virulence factor. Lumpy skin disease virus (LS	-2.17	7.6	-6.88	1.53E-05	6.47
PU09002	Auxin-binding protein ABP20 precursor. Prunus persica (Peach).ABP20. F	-2.06	7.8	-6.78	1.80E-05	6.19
PU01369	expressed protein	-2.05	7.17	-7.38	6.55E-06	7.60
PU09312	Auxin-binding protein ABP20 precursor. Prunus persica (Peach).ABP20. F	-1.99	7.91	-6.38	3.72E-05	5.28
PU10795	expressed protein	-1.96	10.19	-11.40	4.33E-08	15.50
PU29422	ORF124 (Wsv247) (WSSV302). White spot syndrome virus (WSSV).	-1.94	10.57	-12.00	4.33E-08	16.40
PU00843	Floricaula/leafy homolog (PTLF). Populus trichocarpa (Western balsam p(	-1.91	10.38	-11.80	4.33E-08	16.10
PU29442	RING-H2 zinc finger like protein	-1.90	9.76	-10.20	1.18E-07	13.30
PU13110	Floricaula/leafy homolog (PTLF). Populus trichocarpa (Western balsam p(	-1.88	11.05	-11.70	4.33E-08	16.00
PU29512	expressed protein	-1.87	10.89	-11.30	4.33E-08	15.30
PU29991	phospholipase D (AtPLDdelta)	-1.87	9.15	-10.50	8.37E-08	13.80
PU03680	Floricaula/leafy homolog (PTLF). Populus trichocarpa (Western balsam p(	-1.86	10.5	-11.20	4.33E-08	15.10
PU29417	ORFs (Putative primordial protein). Mus musculus (Mouse).	-1.86	10.66	-10.50	8.20E-08	13.90
PU00007	Floricaula/leafy homolog (PTLF). Populus trichocarpa (Western balsam p(	-1.84	8.81	-10.90	5.33E-08	14.70
PU29433	H88. Human herpesvirus 6.H88.	-1.84	9.12	-9.96	1.53E-07	12.90
PU29358	expressed protein	-1.83	10.83	-10.50	8.10E-08	14.00
PU29418	expressed protein	-1.83	10.18	-10.20	1.18E-07	13.30
PU27021	Hypothetical protein XAC0271. Xanthomonas axonopodis (pv. citri).XAC0	-1.82	11.2	-10.20	1.19E-07	13.30

b) Before infection down- regulated.

Table 5.4 Up and down regulated genes in small size uridinia in non-infected samples

The differentially expressed genes can also be plotted onto the genome to see where they are positioned, and to select those that are in the QTL region.

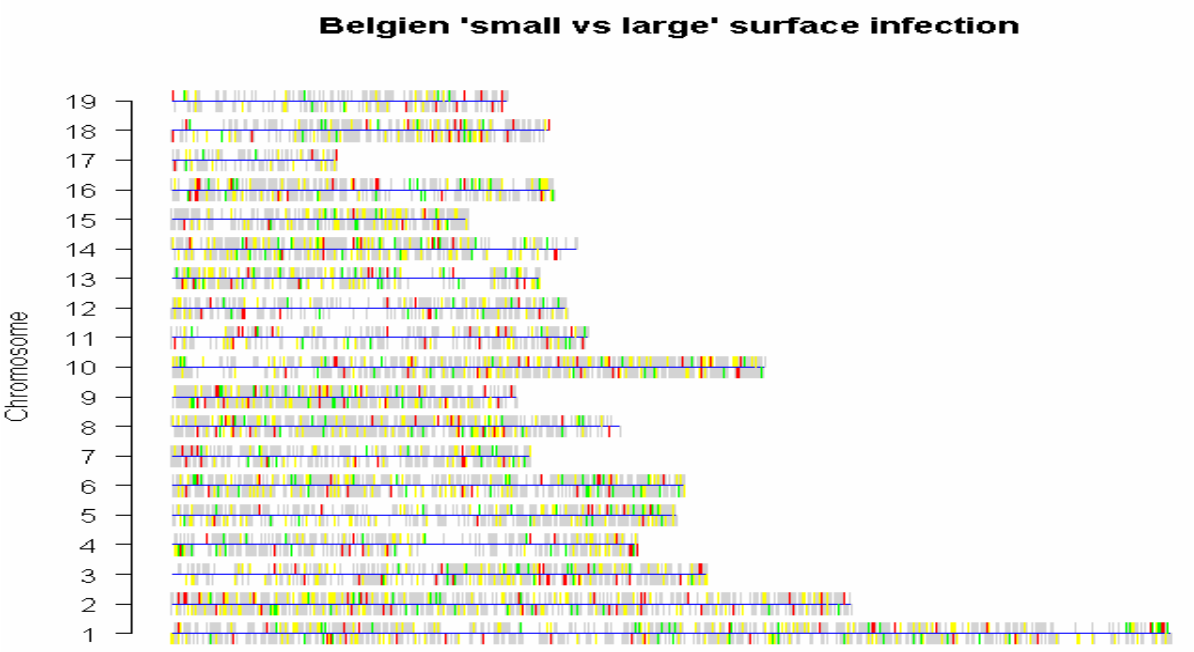


Figure 5.6 Expressed genes plotted onto the genome. (Red) is up-regulated at least 2-fold; (Green) is down-regulated at least 2-fold; (Yellow) is significantly differential but less than 2-fold; (Gray) genes available on the arrays.

Because our aim was to identify genes that were differentially expressed on LG XVI, these were selected. For the results coming from the individual loop, where the individual genotypes are treated as groups, the comparison is done between the L2K and L1K group.

PU.name	Clone	Annotation	M	A	t	P.Value	B
PU12064	I086P27	expressed protein	0.92	8.91	4.99	5.07E-04	1.81
PU23736	S063E03	expressed protein	1.08	7.58	8.93	6.62E-07	10.9
PU13136	F106P20	myosin heavy chain	0.75	7.18	8.37	1.48E-06	9.68
PU01975	A063P72	60S ribosomal protein L27. Pisum sativum (Garden pea).RPL27. POSSIBLY ribosomal protein	0.63	11.05	5.31	2.73E-04	2.6
PU02439	A084P73	HSP like protein	0.62	7.49	4.11	2.91E-03	0.04
PU07022	G094P06	HSP like protein	0.69	7.8	4.5	1.33E-03	0.68
PU20364	M129A01	P0468B07.6 protein. Oryza sativa (japonica cultivar-group).P0468B07.6. POSSIBLY SigA binding protein	1.22	7.87	8.41	1.40E-06	9.8
PU26210	UM49TH02	Beta-1,3 glucanase. Populus x canescens.BGLUC. POSSIBLY glucan endo-1,3-beta-D-glucosidase-like prot	1.99	9.04	6.83	1.64E-05	6.27
PU29234	Q010D20	No annotation avaliable	2.52	7.78	8.36	1.49E-06	9.7
PU29350	Q020G05	Glycosyltransferase (EC 2.4.1.-). Methanosarcina mazei (Methanosarcina frisia).MM1138.	2.52	8.43	7.78	3.45E-06	8.45
PU04824	UB34DPB12	receptor-like serine/threonine kinase, putative	0.96	7.06	6.13	5.80E-05	4.76
PU08400	C024P33	RING-H2 finger protein RHF2a	0.64	9.74	5.96	7.98E-05	4.19
PU00863	A028P61	Ubiquitin. Pisum sativum (Garden pea).PUB2 OR PUB1. POSSIBLY polyubiquitin 4 UBQ4	0.87	10.3	8.57	1.09E-06	10.1
PU22285	P084E12	Myb-like protein (Fragment). Dictyostelium discoideum (Slime mold).MYBA.	0.91	11.81	8.87	7.15E-07	10.7
PU27570	V058G09	ubiquitin/ribosomal protein CEP52	0.88	10.96	10.9	5.82E-08	14.5
PU20613	N016A03	Syntaxin-related protein Nt-syr1. Nicotiana tabacum (Common tobacco).NT-SYR1. POSSIBLY putative synta	0.71	11.2	5.97	7.88E-05	4.21
PU06009	UB59CPH01	EIN2	0.72	8.91	8.09	2.16E-06	9.09
PU20140	M113A06	Phenylalanine ammonia-lyase (EC 4.3.1.5). Populus trichocarpa (Western balsam poplar).PAL. POSSIBLY p	0.93	7.41	5.05	4.45E-04	2.08
PU09192	C074P06	Gastric mucin (Fragment). Sus scrofa (Pig). POSSIBLY GTPase activator protein of Rab-like small GTPases	1.19	10.63	7.96	2.62E-06	8.8
PU24075	T018A08	expressed protein	0.61	7.45	5.36	2.46E-04	3
PU07822	G114P64	Laccase (EC 1.10.3.2) (Fragment). Populus trichocarpa (Western balsam poplar).LAC1. POSSIBLY putative	0.6	7.16	5.34	2.57E-04	2.91
PU11407	I073P94	DNA-binding protein 4 (Fragment). Nicotiana tabacum (Common tobacco).WRKY4. POSSIBLY putative DNA	1.49	9.29	10	1.38E-07	13
PU04367	UB13CPF02	UDP-galactose transporter MSS4	0.63	8.33	5.53	1.78E-04	3.17

Up-regulated genes on LG XVI from the individual loop treated as a group.

PU.name	Clone	Annotation	M	A	t	P.Value	B
PU01200	A039P56	Orf protein. Pisum sativum (Garden pea).ORF. POSSIBLY putative expansin	-0.67	8.19	-5.46	2.01E-04	3
PU26387	UM67TG08	Ribosomal protein 5 precursor. Spinacia oleracea (Spinach).PSRP-5. POSSIBLY ribosomal protein, chloropla	-0.72	10.8	-5.58	1.61E-04	3.28
PU07497	G103P46	expressed protein	-0.87	7.75	-4.72	8.70E-04	1.27
PU09069	C041P20	Ferredoxin-dependent glutamate synthase, chloroplast (EC 1.4.7.1) (Fd-GOGAT).Spinacia oleracea (Spinach	-0.85	11.68	-7.52	5.24E-06	7.83
PU27242	V036F03	expressed protein	-0.62	7.15	-4.35	1.80E-03	0.39
PU04171	UB11CPH12	Unknown protein	-0.95	7.55	-7.03	1.16E-05	6.78
PU04624	UB26CPE09	Unknown protein	-0.71	9.52	-4.69	9.27E-04	1.06
PU06543	G079P74	2-Cys peroxiredoxin. Pisum sativum (Garden pea). POSSIBLY putative 2-cys peroxiredoxin	-0.64	12.49	-9.16	4.64E-07	11.3
PU21409	P012D05	expressed protein	-0.68	8.29	-5.77	1.13E-04	3.81
PU26305	UM59TH07	expressed protein	-0.84	9.19	-7.8	3.34E-06	8.46
PU02877	B008P27	Gamma tonoplast intrinsic protein. Pyrus communis (Pear).PY-GTIP. POSSIBLY putative aquaporin (tonopla	-0.79	8.9	-8.37	1.48E-06	9.7
PU06750	G074P84	Gastric mucin (Fragment). Sus scrofa (Pig). POSSIBLY GTPase activator protein of Rab-like small GTPases	-0.61	8.61	-4.57	1.17E-03	0.76
PU26214	UM51TB11	Nitrogenase iron protein (Fragment). Methanobrevibacter ruminantium.NIFH. POSSIBLY Expressed protein	-0.61	7.83	-5.13	3.85E-04	2.3
PU10800	I040P62	defective in exine formation (DEX1)	-1.02	8.86	-7.34	6.93E-06	7.44
PU00123	A005P70	30S ribosomal protein S31	-1.03	8.24	-6.85	1.60E-05	6.31
PU10795	I045P50	expressed protein	-1.96	10.19	-11.4	4.33E-08	15.5
PU21235	N069G12	CPRD49 protein. Vigna unguiculata (Cowpea).CPRD49.	-0.85	7.14	-5.63	1.46E-04	3.51
PU03613	UA30CPH07	expressed protein	-0.7	8.55	-5.73	1.23E-04	3.66
PU08549	C002P01	elongation factor P (EF-P) like protein	-0.8	7.34	-6.86	1.57E-05	6.43
PU08545	C001P15	Chloroplast ribosomal protein L17. Nicotiana tabacum (Common tobacco). POSSIBLY ribosomal protein L17	-0.85	9.08	-9.59	2.42E-07	12.2
PU09477	C055P47	Chloroplast ribosomal protein L17. Nicotiana tabacum (Common tobacco). POSSIBLY ribosomal protein L17	-0.94	9.41	-6.79	1.76E-05	6.17

Down-regulated genes on LG XVI from the individual loop treated as a group.

The top 20 genes that were highest, respectively least expressed with a high B-statistic value from the direct comparison of L1 and L2 RNA-pools from the samples before and after infection are given below.

PU.name	Annotation - Up-regulated in small size non-infected samples	M	A	t	P.Value	B
PU20185	Bet v I allergen family protein	3.18	8.11	5.800	0.204	0.824
PU10650	chromatin protein family	3.03	8.37	7.560	0.124	2.785
PU12372	nucleotide diphosphate kinase II, chloroplast (NDPK2)	2.85	7.73	7.260	0.204	0.619
PU20165	mitochondrial transcription termination factor family protein / mTERF fam	2.53	10.09	11.700	0.068	3.838
PU09119	glycosyl hydrolase family 17 protein	2.41	8.59	7.840	0.124	2.910
PU20311	alpha-galactosidase, putative / melibiase, putative / alpha-D-galactoside (	2.30	7.78	7.040	0.149	2.018
PU03804	basic endochitinase	2.17	9.36	8.800	0.068	3.655
PU12201	calcineurin B-like protein 10 (CBL10)	2.10	8.16	6.030	0.204	0.794
PU09894	Bet v I allergen family protein	1.97	7.56	5.790	0.214	0.172
PU11342	glutathione S-transferase, putative	1.94	9.48	6.130	0.204	1.284
PU20173	short-chain dehydrogenase/reductase (SDR) family protein	1.92	8.38	8.700	0.068	3.753
PU10056	complex 1 family protein / LVR family protein	1.89	9.99	7.540	0.124	2.766
PU12110	trypsin and protease inhibitor family protein / Kunitz family protein	1.79	9.25	10.000	0.068	4.846
PU03504	glycosyl hydrolase family 29 / alpha-L-fucosidase, putative	1.71	10.27	6.850	0.204	1.634
PU12777	glutathione S-transferase, putative	1.69	9.44	7.150	0.149	2.472
PU22569	drought-responsive family protein non-consensus AT donor splice site at	1.58	8.99	5.650	0.204	0.771
PU03714	Unknown	1.55	11.51	7.170	0.204	1.383
PU03348	hypothetical protein predicted protein, Arabidopsis thaliana	1.38	9.32	5.470	0.214	0.237
PU25160	calcium-transporting ATPase 8, plasma membrane-type / Ca(2+)-ATPase	1.29	9.81	5.690	0.214	0.482
PU09144	ribosomal protein S7 family protein	1.25	9.90	6.280	0.204	1.522

Before infection up-regulated.

PU.name	Annotation - Down-regulated in small size non-infected samples	M	A	t	P.Value	B
PU13518	expressed protein	-2.53	7.91	-6.870	0.164	1.871
PU29698	GCN5-related N-acetyltransferase (GNAT) family protein	-2.05	9.55	-6.450	0.204	0.794
PU29649	UDP-glucose glucosyltransferase, putative	-1.86	8.99	-5.950	0.204	0.755
PU05464	expressed protein	-1.85	10.23	-5.880	0.214	0.279
PU25651	DNA-directed DNA polymerase family protein	-1.79	9.87	-5.980	0.214	0.381
PU12438	expressed protein similar to ABA-inducible protein [Fagus sylvatica] GI:3(	-1.76	9.69	-5.500	0.211	0.590
PU31099	expressed protein similar to ABA-inducible protein [Fagus sylvatica] GI:3(	-1.63	10.47	-5.790	0.204	0.931
PU03204	isocitrate lyase, putative	-1.62	10.25	-6.200	0.214	0.577
PU26988	expressed protein	-1.56	10.36	-6.350	0.211	0.710
PU27490	fructose-1,6-bisphosphatase, putative / D-fructose-1,6-bisphosphate 1-ph	-1.52	8.29	-6.150	0.204	1.303
PU04585	tropinone reductase, putative / tropine dehydrogenase, putative	-1.46	7.17	-5.570	0.204	0.578
PU25797	expressed protein similar to ABA-inducible protein [Fagus sylvatica] GI:3(	-1.46	9.76	-5.120	0.216	0.076
PU25786	expressed protein	-1.44	9.78	-5.790	0.214	0.581
PU27041	expressed protein	-1.42	11.21	-5.750	0.219	0.162
PU21491	expressed protein similar to ABA-inducible protein [Fagus sylvatica] GI:3(	-1.40	10.40	-6.040	0.204	1.246
PU30268	glutathione S-conjugate ABC transporter (MRP2) almost	-1.39	7.24	-5.670	0.204	0.687
PU03677	protein kinase family protein	-1.37	9.52	-5.350	0.214	0.408
PU29396	metal-transporting P-type ATPase, putative (PAA1) nearly	-1.20	9.44	-6.130	0.204	1.349
PU29433	expressed protein similar to ABA-inducible protein [Fagus sylvatica] GI:3(	-1.18	9.15	-6.010	0.204	1.205
PU30467	galactinol synthase, putative	-1.15	9.88	-5.730	0.204	0.857

Before infection down-regulated.

PU.name	Annotation - Up-regulated in small size infected samples	M	A	t	P.Value	B
PU01672	ADP, ATP carrier protein, mitochondrial, putative / ADP/ATP translocase,	8.36	7.84	6.850	0.124	0.007
PU26381	DEIH-box RNA/DNA helicase, putative	5.18	8.54	9.910	0.124	0.225
PU31034	microtubule-associated EB1 family protein	4.83	7.24	6.110	0.124	0.701
PU30672	calmodulin-binding protein-related (PICBP)	3.87	11.89	9.140	0.111	1.467
PU23774	glycine-rich protein / oleosin	3.77	7.76	5.580	0.124	0.093
PU28117	auxilin-related low similarity to SP|Q27974 Auxilin {Bos taurus}	3.57	11.87	10.300	0.111	1.896
PU27927	BSD domain-containing protein	3.38	10.70	9.930	0.111	1.779
PU21572	Unknown	3.27	10.68	9.010	0.111	1.413
PU28136	Unknown	3.21	8.43	6.320	0.124	0.905
PU27946	expressed protein contains Pfam domain PF05003: protein of unknown f(	2.59	7.45	6.880	0.111	1.414
PU11221	vacuolar processing enzyme gamma / gamma-VPE nearly	2.53	9.83	7.060	0.124	0.405
PU27164	ethylene receptor, putative (EIN4)	2.46	8.34	6.090	0.124	0.092
PU03043	wound-responsive protein-related	2.42	9.25	6.010	0.124	0.603
PU10222	hydroxyproline-rich glycoprotein family protein Common family members:	2.39	9.31	6.960	0.124	0.343
PU04962	hypothetical protein	2.29	9.16	5.890	0.128	0.069
PU00163	phosphoribosylformylglycinamidine cyclo-ligase, chloroplast / phosphorib(	2.26	7.16	5.590	0.124	0.176
PU28889	auxin-responsive protein / indoleacetic acid-induced protein 19 (IAA19)	2.24	7.82	5.340	0.124	0.227
PU05193	RNA-dependent RNA polymerase family protein	2.14	7.13	5.860	0.124	0.355
PU10149	chalcone synthase / naringenin-chalcone synthase	2.07	8.54	6.160	0.124	0.757
PU10656	mitochondrial substrate carrier family protein	2.04	9.75	5.470	0.124	0.384

After infection up-regulated.

PU.name	Annotation - Down-regulated in small size infected samples	M	A	t	P.Value	B
PU11889	AP2 domain-containing transcription factor RAP2.4	-4.12	7.76	-5.470	0.124	0.384
PU04563	expressed protein	-3.75	7.29	-8.590	0.111	2.707
PU13072	WD-40 repeat family protein (LEUNIG) contains seven G-protein beta WI	-3.75	7.11	-5.280	0.124	0.124
PU22312	expressed protein similar to PrMC3 [Pinus radiata] GI:5487873	-3.50	7.50	-4.990	0.124	0.057
PU06059	sugar transporter family protein	-3.27	7.11	-5.840	0.119	1.163
PU04737	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein	-3.23	8.28	-5.590	0.124	0.098
PU05518	cellulose synthase, catalytic subunit, putative	-3.12	7.16	-7.150	0.124	0.807
PU00203	alcohol dehydrogenase (ADH)	-3.01	10.55	-8.990	0.111	1.401
PU28775	expressed protein ; expression supported by MPSS	-2.82	8.64	-8.480	0.111	1.931
PU07960	NAD-dependent epimerase/dehydratase family protein	-2.80	7.41	-7.430	0.111	1.869
PU03469	heat shock cognate 70 kDa protein 1 (HSC70-1) (HSP70-1)	-2.61	10.34	-7.710	0.124	0.776
PU22368	fructose-bisphosphate aldolase, putative	-2.58	7.87	-5.570	0.124	0.823
PU07043	expressed protein contains Pfam profile PF04146: YT521-B-like family	-2.52	8.66	-6.070	0.124	0.664
PU04319	expressed protein	-2.52	7.18	-5.230	0.124	0.068
PU03250	metallothionein protein, putative (MT2A)	-2.49	8.57	-6.000	0.124	0.596
PU06937	elongation factor 1-alpha / EF-1-alpha	-2.47	9.45	-6.290	0.111	1.308
PU25301	histone H2A, putative	-2.44	7.69	-6.470	0.124	1.046
PU09457	Gar1 RNA-binding region family protein	-2.43	9.86	-7.100	0.124	0.427
PU07620	6-phosphogluconate dehydrogenase family protein contains Pfam profiles	-2.42	8.17	-6.190	0.111	1.582
PU00702	expressed protein	-2.42	10.91	-7.200	0.124	0.486

After infection down-regulated.

Discussion

Although the genetical genomics approach is a powerful strategy to enrich in candidate genes, and potentially identify the causal gene, one has to realize the assumptions and limitations of our strategy. Firstly, the cause of the QTL effect should be the differential expression of the two alleles of one of the parents, although we have a priori no evidence that this is the case. Secondly, the causal gene should be represented on the microarrays. The POP3 microarray only has approximately 16,000 of the ~40,000 predicted number of genes. Thirdly, the causal gene should be expressed in the tissue that is sampled, in our case leaf disks.

We have identified a number of differentially expressed candidate genes residing within the QTL for uredinia size. In the next phase, the expression of these candidate genes will be examined in genotypes with contrasting uredinia size coming from natural *P. nigra* populations. Three such populations from France have already been analyzed for uredinia size. From one population derived from the river Garonne, 9 clones with small uredinia and 9 clones with large uredinia were sampled. The second population is from the river Loire. From this population 5 clones with large and 5 with small uredinia size were selected. The third population is from the Pyrenees, and similarly 5 clones with large and 5 with small diameter were chosen. Currently, comparative microarrays are being carried out on the six bulks, each made from the extreme phenotypes of each population. These data will tell whether the difference in phenotype is correlated with the differential expression of a candidate gene residing in the QTL for uredinia size. These data are not

available yet. Finally, for candidate genes identified in this way, SNPs will be identified and association mapping carried out in the *P. nigra* association mapping population. The latter population has already been phenotyped for *M. larici-populina* uredinia size.

Genetical genomics on rust in the POP2 pedigree.

Aim

In the POP2 pedigree, two major genomic regions are involved in the control of quantitative resistance to *M. larici-populina* (Dowkiw and Bastien, 2004; Jorge et al., 2005). One of these regions harbours a major resistance factor, R_{US}, that controls uredinia size; and the second region corresponds to a *Populus deltoides* inherited qualitative resistance gene (R_1) that was shown to have a residual effect on different components of the quantitative resistance (see results for disease and pest traits and QTL discovery section). In order to determine the genetic basis of such resistance in poplar, we have set up a genomics approach that aims in the first place at identifying a large array of differentially expressed genes in POP2 genotypes differing by the presence/absence of R_1 and/or R_{US}. Selected molecular markers that were of particular interest based on the scientific literature, and colocation with QTLs already mapped in the POP2 pedigree will be tested for their expression in hundreds of POP2 genotypes in order to map rust-resistance-related eQTL.

Experimental approach

The 2 POP2 genitors and four F_1 genotypes differing by the presence/absence of R_1 and/or R_{US} were selected as extreme individuals to test their gene expression levels when challenged by rust. Previous transcriptomics studies on qualitative resistance have shown that gene expression differences could not be detected in poplar leaves challenged by rust before 48 hours post-inoculation (hpi) (Duplessis and Martin; unpublished data). In the present study, leaves from the various selected genotypes were inoculated with the *M. larici-populina* 98AG69 strain (see disease trait section). Leaves were maintained in controlled conditions and sampling was performed at 48 and 96 hpi to compare gene expression during time-course infection by the fungus. Ribonucleic acids isolated from leaf samples were labelled and hybridized with poplar PICME 28K cDNA microarrrays (see genomic technologies section). Candidate genes were

selected from duplicated experiments on the basis of their expression levels at 96 hpi compared to 48 hpi.

Results

In the present approach, we identified several dozens of candidate genes that showed important differences in gene expression during time-course infection by rust. Careful attention was given to genes that showed contrasting patterns of expression between extreme genotypes selected in the POP2 progeny for differences in uredinia size and between the POP2 genitors. Several genes corresponded to typical targets activated in the context of qualitative resistance. These genes encode several pathogenesis-related proteins such as PR-1, PR-5 or PR-10. Transcript levels for such genes were higher in leaves of the genotypes that showed smaller uredinia size (R_{US}). Differential gene expression between contrasting parents and contrasting genotypes was confirmed by real-time quantitative RT-PCR at 48 and 96 hpi during time-course infection by rust. Measurement of expression levels by RT-PCR was tested on triplicates of experiment in order to test reproducibility of the technique prior to further use in the genetical genomics approach.

Discussion

Quantitative resistance to rust in the POP2 pedigree is a complex trait that showed differential expression of typical defence-related genes involved in resistance to pathogen attack in plant. Our results tend to indicate that these functions may participate to the quantitative response to *M. larici-populina* in poplar. Our results do not allow at present to determine whether this observation is due to a residual effect of R_1 in POP2 genotypes or to the specific activation of cellular functions through the *Rus* factor. We plan to investigate the expression profiles of selected candidate genes in larger subsets of the pedigree in order to map expression-QTL related to quantitative resistance. Such approach should give us some insight in the programs activated by R_{US}. Selection of candidate genes on the basis of expression studies is of great interest but we may miss key cellular components such as signalling components that act upstream of the selected candidate genes. Several genes falling in genomic regions covered by QTLs for quantitative resistance to rust (see results for disease traits and pest and QTL discovery section)

in the *P. trichocarpa* genome sequence may also be related to pest-resistance (e.g., receptor-like kinases). Investigation of eQTL of such candidates will also be performed in the POP2 pedigree.

5.4 Conclusions

We have shown that the combination of genetics and genomics provides an outstanding opportunity to elucidate the genes underling QTL. In POPYOMICS for the first time in *Populus* we made considerable progress in elucidating genes underlying adaptation to drought and ozone and differential gene expression associated with QTL for resistance and tolerance to the *Melampsora larici-populina* rust. Using the natural *P. nigra* population and other approaches it will in future be possible to provide proof of concept that the candidates identified are associated with traits of interest.

5.5 References

1. Kirst M, Myburg AA, De Léon JPG, Kirst ME, Scott J, Sederoff R (2004) Coordinated genetic regulation of growth and lignin revealed by quantitative trait locus analysis of cDNA microarray data in an interspecific backcross of eucalyptus. *Plant Physiol*, 135, 2368-2378
2. Plomion C, Bahrman N, Costa P, Dubos C, Frigerio J-M, Gion J-M, Lalanne C, Madur D, Pionneau C, Gerber S (2004) Proteomics for genetic and physiological studies in forest trees: application in maritime pine. In *Molecular Genetics and Breeding of Forest Trees*. Edited by Kumar S, Fladung M. New York: Food Products Press, 53-79
3. Neale DB, Savolainen O (2004) Association genetics of complex traits in conifers. *Trends Plant Sci,* 9, 325-330
4. Sterky F, Bhalerao RR, Unneberg P, et al. (2004) A Populus EST resource for plant functional genomics. *Proceedings of the National Academy of Sciences of the United States of America,* 101(38), 13951-13956
5. Sjodin A, Bylesjo M, Skogstrom O, et al. (2006) UPSC-BASE - Populus transcriptomics online. *Plant Journal*, 48(5), 806-817

6. FinlaysonPitts BJ, Pitts JN (1997) Tropospheric air pollution: Ozone, airborne toxics, polycyclic aromatic hydrocarbons, and particles. *Science*, 276(5315), 1045-1052

7. Tuskan GA, DiFazio S, Jansson S, et al. (2006) The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). *Science*, 313(5793), 1596-1604

8. Fowler D, Cape JN, Coyle M, et al (1999) The global exposure of forests to air pollutants. *Water Air and Soil Pollution,* 116(1-2), 5-32

9. Miller JD, Arteca RN, Pell EJ (1999) Senescence-associated gene expression during ozone-induced leaf senescence in Arabidopsis. *Plant Physiology,* 120, 1015-1023

10. Karnosky DF, Gagnon ZE, Dickson RE, Coleman MD, Lee EH, Isebrands JG (1996) Changes in growth, leaf abscission, and biomass associated with seasonal tropospheric ozone exposures of Populus tremuloides clones and seedlings. *Canadian Journal of Forest Research-Revue* Canadienne De Recherche Forestiere, 26, 23-37

11. Fry SC, Smith RC, Renwick KF, Martin DJ, Hodge SK, Matthews KJ (1992) Xyloglucan Endotransglycosylase, a new wall-loosening enzyme-activity from plants. *Biochemical Journal,* 282, 821-828

12. Dowkiw A, Bastien C (2004) Characterization of two major genetic factors controlling quantitative resistance to *Melampsora larici-populina* leaf rust in hybrid poplars: strain-specificity, field expression, combined effects, and relationship with a defeated qualitative resistance gene. *Phytopathology,* 94, 1358-1367

13. Jorge V., Dowkiw A., Faivre-Rampant P., Bastien C. (2005) Genetic architecture of qualitative and quantitative *Melampsora larici-populina* leaf rust resistance in hybrid poplar: genetic mapping and QTL detection. *The New Phytologist,* 167, 113-127

14. Pandey S, Ranade SA, Nagar PK, Kumar N (2000) Role of polyamines and ethylene as modulators of plant senescence. *Journal of Biosciences*, 25, 291-299

15. Pastori GM, Foyer CH (2002) Common components, networks, and pathways of cross-tolerance to stress. The central role of "redox" and abscisic acid-mediated controls. *Plant Physiology*, 129, 460-468

16. Clarke BB, Greenhalghweidman B, Brennan EG (1990) An assessment of the impact of ambient ozone on field-grown crops in New-Jersey using the EDU method .1. White potato (Solanum-Tuberosum). *Environmental Pollution*, 66(4), 351-360

17. Borevitz JO, Maloof JN, Lutes J, et al. (2002) Quantitative trait loci controlling light and hormone response in two accessions of Arabidopsis thaliana. *Genetics*, 160(2), 683-696

18. Marie BA, Ormrod DP (1986) Dose-response relationships of the growth and injury effects of ozone and sulfur-dioxide on brassicaceae seedlings. *Canadian Journal Of Plant Science,* 66(3), 659-667

19. Rowland-Bamford AJ, Borland AM, Lea PJ, Mansfield TA (1989) The role of Arginine Decarboxylase in modulating the sensitivity of barley to ozone. *Environmental Pollution*, 61, 95-106

20. Legocka J, Kluk A (2005) Effect of salt and osmotic stress on changes in polyamine content and arginine decarboxylase activity in Lupinus luteus seedlings. *Journal of Plant Physiology*, 162, 662-668

21. Hao YJ, Kitashiba H, Honda C, Nada K, Moriguchi T (2005) Expression of arginine decarboxylase and ornithine decarboxylase genes in apple cells and stressed shoots. *Journal of Experimental Botany*, 56, 1105-1115

6. POPYOMICS Expoitation and dissemination of results

General Comments

The POPYOMICS project was accompanied by a website – www.soton.ac.uk/~popyomic. In 2007 this website is being updated and will in future house the publicly available data from POPYOMICS, once that is published. This will be an important resource for the scientific community and will include yield and disease traits, in a relational database, experimental descriptions and annotations, descripton of triats, QTL mapping output and graphical display of QTL. This is being undertaken as part of the EVOLTREE (www.evoltree.org) F6 netowrk of excellence in which several POPYOMICS partners are participating.

The POPYOMICS project developed a leaflet that was widely distributed across a large number of scientific meetings. In addition, an annual stakeholder meeting was held in conjunction with the energy biomass project in the UK 'BEGIN' where up to fifity growers, breeders, technologist and energy specialist were present to hear about the latest findings that were presented by G. Taylor each year.

International relationships were developed with the USA, including the *Populus* genome sequencing effort and other large scale poplar genomics projects particularly in Canada and the USA. POPYOMICS was listed on the IPGC (International Poplar Genome Consortium) site and the experimental facilities were open for other researchers to undertake experiments, as agreed by the Project Committee. Several international collaborations were developed in this way.

A large number of presentations were made at scientific meetings and these are listed below. POPYOMICS scientists were energetic in disseminating the results of the project widely. Simialrly a very large number of refereed publications have been forthcoming or are in progress. These are all listed below.

1. Refereed journal papers / Papers published relevant to POPYOMICS

Beritognolo I., Sabatti M., Brosché M., G Scarascia Mugnozza G. (2007). Functional genomics to discover genes for salt tolerance in annual and perennial plants. In: Abdelly C., Ozturk M., Grignon C. (Eds) "Biosaline Agriculture and Salinity Tolerance in Plants". Birkhäuser Verlag AG. Basel. In press

Bunn SM, Rae AM, Herbert CS, Taylor G. (2004) Leaf level productivity in Populus grown in short rotation coppice for biomass energy. Forestry, 77, 307-323

Bastien C, Jorge V, Dowkiw A, et al. (2004) Genetic improvement of poplars
Biofutures, 247, 33-37

Dillen SY, Marron N, Koch B, Ceulemans R (2007) Genetic variation in stomatal traits and carbon isotope discrimination in two hybrid poplar families, in preparation

Dillen SY, Storme V, Marron N, Bastien C, Pinel MPC, Ceulemans R, Boerjan W (2007) Genotype by environment interactions in two related poplar families grown in contrasting sites across Europe. 1. Tree dimensions at the end of the second growing season, in preparation

Dillen SY, Marron N, Bastien C, Ricciotti L, Salani F, Sabatti M, Pinel MPC, Rae AM, Taylor G, Ceulemans R (2007) Effects of environment and progeny on biomass estimations of five hybrid poplar families grown at three contrasting sites. For Ecol Manage, in press

Dowkiw A, Rae AM, Jorge V, Bastien C, Taylor G. (2007) Genetic architecture of quantitative Melampsora larici-populina leaf rust resistance in hybrid poplar, in preparation

Dowkiw A, Bastien C (2007) Presence of defeated qualitative resistance genes frequently has major impact on quantitative resistance to Melampsora larici-populina leaf rust in P.xinteramericana hybrid poplars. Tree Genetics and Genomes, 3, 261-274

Dowkiw A, Bastien C (2005) Characterization of two major genetic factors controlling quantitative resistance to Melampsora larici-populina leaf rust in hybrid poplars: Strain specificity, field expression, combined effects, and relationship with a defeated qualitative resistance gene.
Phytopathology, 95, 124-124

Dowkiw A, Bastien C (2004) Characterization of two major genetic factors controlling quantitative resistance to Melampsora larici-populina leaf rust in hybrid poplars: Strain specificity, field expression, combined effects, and relationship with a defeated qualitative resistance gene, Phytopathology, 94, 1358-1367

Ferris R, Long L, Bunn SM, Bradshaw HD, Rae AM, Taylor G. (2002) Leaf stomatal density and index: the identification of putative QTL in relation to elevated CO_2 in poplar. Tree Physiology, 22, 633-640.

Gaudet,M., Jorge,V., Paolucci,I., Beritognolo,I., Scarascia Mugnozza,G., and Sabatti,M. (2007) Genetic linkage maps of *Populus nigra* L. including AFLPs, SSRs, SNPs, and sex trait. Tree Genetics and Genomes DOI: 10.1007/s11295-007-0085-1.

Gaudet,M., Fara,A.G., Sabatti,M., Kuzminsky,E., and Mugnozza,G. (2007) Single-reaction for SNP Genotyping on Agarose Gel by Allele-specific PCR in Black Poplar (*Populus nigra* L.). Plant Molecular Biology Reporter DOI: 10.1007/s11105-007-0003-6.

Jansson S, Douglas CJ (2007) *Populus*: A model system for plant biology. Annu. Rev. Plant Biol.: 2007 58:435-458

Jorge V, Dowkiw A, Faivre-Rampant P, et al. (2005) Genetic architecture of qualitative and quantitative Melampsora larici-populina leaf rust resistance in hybrid poplar: genetic mapping and QTL detection New Phytologist, 167, 113-127

Kohler A, Rinaldi C, Duplessis S, Baucher M, Geelen D, Hellsten U, Duchaussoy F, Difazio S, Boerjan W & Martin F. Genome-Wide identification of *NBS* resistance genes in *Populus trichocarpa*. Submitted to Plant Molecular Biology (in revision)

Marron N, Dillen SY, Ceulemans R (2007) Leaf determinants of productivity in poplar depend on environmental conditions and genetic background. Environ Exp Bot, 61, 103-116

Marron N & Ceulemans R (2006) Genetic variation of leaf traits related to productivity in a *Populus deltoides x P. nigra* family. Can J For Res 36, 390-400

Marron N, Bastien C, Sabatti M, Taylor G & Ceulemans R (2006) Plasticity of growth and sylleptic branchiness in two poplar families grown at three sites across Europe. Tree Physiol 26, 935-946

Marron N, Storme V, Dillen SY, Bastien C, Ricciotti L, Salani F, Pinel MPC, Ceulemans R, Boerjan W (2007) Genotype by environment interactions in two related poplar families grown in contrasting sites across Europe. 2. Aboveground wood production at the end of the second growing season, in preparation

Marron N, Sabatti M, Ricciotti L, Salani F, Bastien C, Beritognolo I, Gaudet M, Paolucci I, Dillen SY, Pinel M, Taylor G, Scarascia Mugnozza G. (2007) Influence of environment on an intraspecific P. alba family grown at three sites across Europe. I. Plasticity of growth and biomass production over three growing seasons, in preparation.

Rae AM, Pinel MPC, Bastien C, Sabatti M, Street N, Tucker J, Dixon C, Marron N, Dillen SY, Taylor G (2007) QTL for yield in bioenergy *Populus*: Identifying G x E interactions from growth at three contrasting sites. Tree Gene Gen, in press

Rae AM, Ferris R, Bunn SM, Taylor G (2006) Elucidating QTL for growth and development in elevated CO_2. Plant, Cell and Environment, 29, 1730-1741

Rae, Am, Tricker Pj, Bunn Sm And Taylor G (2007) Adaptation of tree growth to elevated CO_2: QTL for biomass in *Populus.* New Phytologist, 175, 59-69

Rae AM, Robison KM, Street NR, Taylor G (2004) Morphological and Physiological Traits Influencing Biomass Productivity in Short Rotation Coppice Poplar. Canadian Journal of Forest Research 34, 1488-1498

Rinaldi C, Kohler A, Frey P, Duchaussoy F, Ningre N, Couloux A, Wincker P, Le Thiec D, Fluch S, Martin F & Duplessis S (2007) Transcript profiling of poplar leaves upon infection with compatible and incompatible strains of the foliar rust *Melampsora larici-populina*. Plant Physiology, 144, 347-366

Rodriguez-Acosta M, Rae Am, Street Nr, Trewin H, Taylor G (2007-). Leaf and stomatal development in *Populus* in relation to environment, in preparation

Rohde, A., Ruttink, T., Hostyn, V., Sterck, L., Van Driessche, K., Boerjan, W. Dynamic gene expression changes during the induction and maintenance of dormancy in poplar apical buds. J. Exp. Botany, in revision

Rohde, A., Bhalerao, R.P. (2007) Plant dormancy in a perennial context. Invited review, Trends in Plant Science 12, 217-223

Ruttink, T., Arend, M., Morreel, K., Storme, V., Rombauts, S., Bhalerao, R., Boerjan, W., Rohde, A. (2007) A molecular timetable for apical bud formation and dormancy induction in poplar. Plant Cell 19, XXX-XXX

Segerman B, Jansson S, Karlsson J (2007) Characterization of genes with tissue-specific differential expression patterns in Populus. Tree Genetics & Genomes, 2007, on line publication

Sims REH, Hastings Schlamadinger AB, Taylor G & Smith P (2006) Energy Crops: Current Status and Future Prospects Global Change Biology, 12, 2459-2471

Sjodin A, Bylesjo M, Skogstrom O, Eriksson D, Nilsson P, Ryden P, Jansson S, Karlsson J (2006) UPSC-BASE - Populus transcriptomics online. Plant Journal, 2006 48, 806-817

Andreas Sjödin, Kirsten Wissel, Max Bylesjö, Oskar Skogström, Henrik Antti, Johan Trygg, Stefan Jansson (2007) Global expression profiling in leaves of free-growing aspen, in preparation

Skogström O, Street NR, Hyoshin L, Taylor G and Jansson S. A genetical genomics approach to understand drought tolerance in *Populus*, in preparation

Street NR And Tucker J, Broesche M, Jangasjaarvi J, Broadmeadow MB, Taylor G (2007-). Insights into the genetic architecture of ozone response in *Populus*, in preparation

Street Nr , Skogstrom O, Sjodin A, Tucker J, Rodriguez-Acosta M, Nilsson P, Jansson S, Taylor G. (2006) The genetics and genomics of drought response in *Populus*. Plant Journal, 48, 321-341

Taylor G. Tallis MJ, Giardina CP, Kevin Ep, f Miglietta, Pooja S. Gupta, Beniamino Gioli, Carlo Calfapietra, Birgit Gielen, Mark E. Kubiske, Giuseppe E. Scarascia-Mugnozza, Katre Kets, Stephen P. Long and David F. Karnosky (2007) Future atmospheric CO_2 leads to delayed autumnal senescence in *Populus* over two continents Global Change Biology, in press

Taylor G, Beckett KP, Bunn SM, Robinson KM, Rae AM. (2001) Identifying QTL for yield in UK biomass poplar. Aspects of Applied Biology, 65, 173-182

Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen GL, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroeve S, Dejardin A, Depamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjarvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leple JC, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouze P, Ryaboy D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai CJ, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, de Peer YV, Rokhsar D (2006) The genome of black cottonwood, Populus trichocarpa (Torr. & Gray) Science, 2006, 313, 1596-1604

ZAINA G., MORGANTE M. (2004) Nucleotide diversity and linkage disequilibrium in *Populus nigra*, in preparation

2. Abstracts of oral presentations and dissemination through seminars

Beritognolo I., Sabatti M., Loreto F., Brilli F., Salani F., Pinelli P., Kuzminsky E., Kangasjärvi J., Scarascia Mugnozza G. e Valentini R. (2006) Ecophysiological response and transcriptome regulation in mediterranean *Populus alba* genotypes submitted to shock salinity stress. International Conference on Biosaline Agriculture & High Salinity Tolerance. Tunis, Tunisia
Beritognolo I., Sabatti M., Gaudet M., Paolucci, I. Kuzminsky E., Scarascia Mugnozza G., Brilli F., Loreto F., and Valentini R. (2007) Comparative transcriptomics to understand variability of salinity tolerance in white poplar. 51st Annual Congress of the Italian Society of Agricultural Genetics (SIGA). Riva del Garda (TN), Italy. September 23rd – 26th

Beritognolo I., Brilli F., Gaudet M., Kuzminsky E., Loreto F., Paolucci I., Sabatti M., Salani F., Scarascia Mugnozza G., and Valentini R. (2007) Molecular and ecophysiological approach to

identify genes controlling salinity tolerance in Populus *alba L. Sixth National Congress of the Italian Society of Silviculture and Forest Ecology (SISEF). Arezzo, Italy. September 25th – 27th*

Dillen SY, Marron N, Ceulemans R (2006) Leaf determinants of productivity in poplar depend on environmental conditions and genetic background. Programme and Conference Abstract Book pp. 292-293. COST Action E-28 Final meeting, 5-6 October, Madrid, Spain

Dillen SY, Marron N, Al Afas N, Laureysens I, Ceulemans R (2006) Experience and research on short rotation in Flanders. Biomass for Energy: Challenges for Agriculture. Conference Book pp. 80-81. North Sea Bioenergy, 25-26 September, Bruges, Belgium

DUPLESSIS S, KOHLER A, RINALDI C, FREY P & MARTIN F (2004) Characterization of the early response of poplar to rust infection using expression profiling. *12th New Phytologist Symposium on Functional Genomics of Environmental Adaptation in* Populus. Gatlinburg TE, USA

Jansson, S (2003) SEB Meeting, Southampton, England

Jansson , S (2003) IUFRO Tree Biotechnology conference Umeå, Sweden

Jansson S (2003) SPPS Congress, Rönne, Bornholm, Denmark

Jansson S (2003) XV Genome Sequence Analysis Congress, Savannah, USA

Jansson S (2004) Plant Genomics: technological innovation, natural diversity, risks and consumer worries. ESOF conference, Stockholm, Sweden

Jansson S (2004) 3rd Plant Genomics European Meeting (GEM), Lyon, France

Jansson S (2005) 2nd ESF Functional genomics and disease FUGE conference Satellite meeting, Oslo, Norway

Jansson S (2006) 3rd EPSO conference, Visegrad, Hungary

Jansson S (2006) 5th Plant Genomics European Meeting (GEM), Venice, Italy

Marron N, Dillen SY, Bastien C, Poursat P, Gauvin J, Rae AM, Pinel MPC, Taylor G, Sabatti M, Ricciotti L, Salani F, Ceulemans R (2006) Site conditions affect expression of genetic variability for biomass production and its determinants in five poplar families at three sites across Europe. 1. Direct estimation of growth performance: tree dimensions and biomass production. Abstract Book pp. 86-87. Fourth International Poplar Symposium, IUFRO, 5-9 June, Nanjing, China

Marron N, Dillen SY, Ceulemans R (2005) Heterosis and genotype x environment interaction in biomass production of two poplar families grown at two contrasting sites in Europe. 14th European Biomass Conference and Exhibition, ETA Florence, 17-21 October, Paris, France.

Marron N, Monclus R, Dillen SY, Ceulemans R, Dreyer E, Brignolas F (2007) Diversity of leaf traits related to productivity and water-use efficiency in poplars. Refining plant functional classifications for earth system modelling. QUEST PFT-FTI workshop, 7-9 February, Alicante, Spain

Morgante M and Zaina G (2004) Sequence variation and linkage disequilibrium in *Populus* . 12th new Phytologist Symposium on Functional Genomics of Environmental Adaptation in *Populus*, Gatlinburg, TN, USA. – 10-13

Morgante M. (2004) Advanced mapping tools in plant genetics. Agricultural Biotechnology International (ABIC) Conference Cologne, Germany – 12-15 September 2004

Morgante M (2004) Analysis of Tree Genome Structure, Diversity and their relationship.14th Federation of European Societies of Plant Biology Congress, Cracow, Poland

Morgante M (2004) Sequence diversity and linkage disequilibrium in perennial outcrossers ITMI/BBSRC Linkage Disequilibrium Workshop, Novotel Barossa Valley Resort, South Australia – 4-6 April

Morgante M (2004) 6th FISV CONGRESS, Riva del Garda (Italy) Population genetics and Evolution Nucleotide diversity and linkage disequilibrium in *Populus nigra* – 30th September / 3rd October 2004 Minisimposium

Rae AM, Sewell MM, Robinson KM, & Taylor G. (2003) Energy from trees? QTL discovery in biomass poplar grown in short rotation. Paper presented at the Society of Experimental Biology Annual General meeting. Southampton UK 31st Mar-4th

Rohde A, Ruttink T, Dillen SY, Marron N, Fabbrini F, Storme V, Jorge V, Rae AM, Paolucci I, Gaudet M, Taylor G, Ceulemans R, Steenackers M, Sabatti M, Bastien C, Boerjan W (2006) An integrated approach to bud set in poplar: phenotypes, candidate genes, and QTLs. Programme and Conference Abstract Book pp. 92-93. COST Action E-28 Final meeting, 5-6 October, Madrid, Spain

Rohde, A. (2007) From phenotypes to genes: an integrated approach to bud dormancy in poplar. 8th Plant Cold Hardiness Seminar. Invited keynote speaker, Saskatoon, Canada

Rohde, A. (2004) Towards a transcript map for bud dormancy in poplar. Invited speaker, 3rd SPSS PhD student Congress “Progress in Plant Biology”, Oslo, Norway

Rohde, A., Ruttink, T., Storme, V., Jorge, V., Paolucci, I., Gaudet, M., Sabatti, M., Bastien, C., Boerjan, W. An integrated approach to bud set in poplar. IUFRO Tree Biotechnology 2007, Ponta Delgada, Portugal, 03.-08.06.2007

Rohde, A., Ruttink, T., Dillen, S., Marron, N., Fabbrini, F., Storme, V., Jorge, V., Rae, A., Paolucci, I., Gaudet, M., Taylor, G., Ceulemans, R., Steenackers, M., Sabatti, M., Bastien, C. Boerjan, W. (2006) An integrated approach to bud set in poplar: phenotypes, candidate genes,

and QTLs. IUFRO Conference “Population Genetics and Genomics of Forest Trees: From gene function to evolutionary dynamics and conservation”. Madrid, Spain

Rohde, A., Boerjan, W. (2004) Deciphering the transcriptome during bud development and dormancy in poplar. Conférence Jacques Monod “Integrative biology: dissecting cross talk between plant signalling pathways”, Roscoff, France

Ruttink, T., Storme, V., Druart, N., Bhalerao, R., Boerjan, W., Rohde, A. (2006) Short-day induced bud set in poplar: regulators and processes. Plant and Microbe Adaptations to Cold. Salsomaggiore, Italie

Ruttink, T., Storme, V., Druart, N., Bhalerao, R., Boerjan, W., and Rohde, A. (2006) The dormancy transcriptome in apical buds of poplar. COST E28 meeting on Functional genomics of tree maturation and reproduction. Invited speaker, Madrid, Spain

Ruttink, T., Storme, V., Druart, N., Bhalerao, R., Boerjan, W., Rohde, A. (2005) The dormancy transcriptome in apical buds of poplar. IUFRO Tree Biotechnology 2005, Pretoria, South Africa

Sabatti M., Gaudet M., Paolucci I., Beritognolo I., Ricciotti L., Salani F., Fabbrini F., Baldasso M., Bastien C., Marron N., Dillen S., Taylor G., Rohde A., Scarascia Mugnozza G. (2007) Phenotipic plasticity of adaptive traits in Populus spp. Sixth National Congress of the Italian Society of Silviculture and Forest Ecology (SISEF). Arezzo, Italy

Scarascia Mugnozza G, Sabatti M, Kuzminsky E, Massacci A, Beritognolo I, Paolucci I, Gaudet M, Pietrini F, Paris P, Loreto F (2007) Molecular bases of adaptive traits in Mediterranean poplars: potentials for biomass production improvement and environmental amelioration. Workshop on Ecophysiology, Biology and Genetics of Poplars. Florence, Italy

Street NR, Rae AM, Trewin H, Taylor G (2006) Poplar Genomics for Biomass Production. ESF-Welcome Trust: Crop Genomics, Trait Analysis and Breeding, Cambridge, UK

Street NR, Tucker J, Skogström O, Brosche M, Sjödin A, Rodríguez-Acosta M, Kangasjärvi J, Jansson S, Taylor G (2006) A Genetical-genomics approach to understanding abiotic stress response in *Populus*. IUFRO/COST-E28 Population Genetics and Genomics of Forest Trees, Madrid, Spain

Street NR, Tucker J, Skogström O, Brosche M, Sjödin A, Rodríguez-Acosta M, Kangasjärvi J, Jansson S, Taylor G (2006) A genetical-genomics approach to understanding abiotic stress response in *Populus*. Plant Gems5, Venice, Italy

Taylor G, Tallis MJ, Street NR, Rae AM, Graham LE, Tucker J and Karnosky DF (2006) Genetical genomics and response of poplar to a changing environment, International poplar symposium, Nanjing, China

Taylor G. (2005) Biotechnology for Bioenergy – can we make the perfect tree? Invited speaker in session entitled ‘Is there a future for energy crops?’BA Festival of Science, Dublin, Ireland

Taylor G. (2005) Invited to a closed meeting between Department of Energy, USA and BP to consider 'Plant Genomics for Biofuels', talk presented: 'Optimised Yield For Bioenergy – A Systems Biology approach? Washington, USA

Taylor G, Tallis MJ, Tricker PJ, Street NR, Rae AM, Jansson S, Graham LE (2004) GEMS 3. Plant Genomic European Meeting. Invited speaker, 'Genomic approaches to understand adaptation of trees to increasing CO_2', Lyon, France

Taylor G. (2004) Poplar as a model tree for understanding the responses to global climate change, International symposium on *Populus*, Goettingen, Germany

Taylor G, Graham LE, Street NR, Jansson S and Tucker J. (2004) Air Pollution workshop symposium 'Poplar Functional Genomics and Global Change – Using Microarrays to Unravel Responses to CO_2 and Ozone, Rhinelander, USA

Taylor G. (2003). Tree Genes and Global Climate Change, ECOGENE International project meeting, Versailles, France

Taylor G, P.J. Tricker, L. Graham, N.R. Street, A.M. Rae, M.J. Tallis, S. Jansson (2004) 12th Tree Genes and adaptation to global change, International New Phytologist Symposium, Smokey Mountains, Tennessee

G. Taylor, P.J. Tricker, L. Graham, N.R. Street, A.M. Rae, M.J. Tallis, S. Jansson (2005) The Genetic Basis of Plant Adaptation to Elevated CO2. International Botanical Congress, Vienna, Austria

3. Posters presented at conferences

Beritognolo I., Loreto F., Sabatti M., Salani F., Pinelli P., Brilli F., Kuzminsky E., Scarascia Mugnozza G (2005) Caratterizzazione della risposta di breve termine allo stress salino in due genotipi di *Populus alba L*. Fifth Sixth National Congress of the Italian Society of Silviculture and Forest Ecology (SISEF). Grugliasco (TO), Italy

Beritognolo I., Sabatti M., Loreto F., Brilli F., Salani F., Pinelli P., Kuzminsky E., Kangasjärvi J., e Scarascia Mugnozza G (2006) Short term ecophysiological response and transcriptome regulation in *Populus alba* genotypes under salinity stress. International Poplar Symposium. Nanjing, China

Dillen SY, Marron N, Bastien C, Ricciotti L, Salani F, Sabatti M, Pinel MPC, Rae AM, Taylor G, Ceulemans R (2007) Effects of environment and progeny on biomass estimations of five hybrid poplar families grown at three contrasting sites. 15th European Biomass Conference and Exhibition, Berlin, Germany

Dillen SY, Marron N, Bastien C, Poursat P, Gauvin J, Sabatti M, Ricciotti L, Salani F, Rae AM, Pinel MPC, Street NR, Taylor G, Ceulemans R (2006) Site conditions affect expression of genetic variability for biomass production and its determinants in five poplar families at three sites across Europe. 2. Indirect estimation of growth performances: ramification and phenology. Abstract Book pp. 88-89. Fourth International Poplar Symposium, Nanjing, China

Dillen SY, Marron N, Ceulemans R (2005) Environmental influence on the aboveground biomass production of two poplar families grown in Central France and Northern Italy. 14th European Biomass Conference and Exhibition, Paris, France

Duplessis S, Kohler A, Rinaldi C, Frey P & Martin F (2004) Dissecting the early molecular events in poplar leaves upon interaction with rust. *XL annual meeting of the Argentine Society for Biochemistry and Molecular Biology.* Iguazu, Missiones, Argentina

Gaudet M, Ricciotti L, Paolucci I, Sabatti M, Scarascia Mugnozza G (2005). Mappa genetica di un incrocio intraspecifico di pioppo nero (Populus nigra L.) basata su marcatori molecolari SSR e AFLP. V Congresso Nazionale SISEF, Grugliasco (TO) Italia, 27-30 Settembre 2005. Poster vincitore del premio "Umberto Bagnaresi"

Gaudet M, Paolucci I, Jorge V, Sabatti M, Scarascia Mugnozza G (2006). Genetic map of black (Populus nigra L.) intra-specific cross based on AFLP, SSR, and SNP markers. Fourth International Poplar Symposium, Nanjing, China

Marron N, Ceulemans R (2004) Heterosis and genotype x environment interaction of two poplar pedigrees grown in three contrasting environments. 12th New Phytologist Symposium. Functional genomics of environmental adaptation in *Populus*, Gatlinburg, Tennessee, USA

Morgante M and Zaina, G (2004) Nucleotide diversity and linkage disequilibrium in *Populus nigra XLVIII ANNUAL CONGRESS SIGA*, Lecce (Italy) – 15th / 18th September (poster C.10)

Paolucci I., Gaudet M., Sabatti M., Muleo R., Terzoli S., Scarascia Mugnozza G. (2006) Intraspecific genetic linkage maps of *Populus alba* L. using AFLP, SSR and RAPD markers. Fourth International Poplar Symposium, Nanjing, China

Pinel, M.P.C., Rae AM, Tuskan, G.A. and Taylor G. Branching Out Early (2004) The initiation of Sylleptic Branching in Populus. Poster presented at the 12th New Phytologist Symposium. Gatlinburg Tennessee, USA

Pereda V, Kohler A & Martin F (2004) Structure and expression profile of the Pi transporters PHO1, Pht1, Pht2 and Pht3 gene families in *Populus* indicate a broad role in Pi homeostasis. *12th New Phytologist Symposium on Functional Genomics of Environmental Adaptation in* Populus. Gatlinburg Tennessee, USA

Rae AM, Pinel MPC, Bastien C, Sabbatti M, Trewin HR, Street NR, Tucker J, Dixon CA, Taylor G (2006) Identifying GxE interactions influencing growth characteristics and QTL

discovery for an F2 population of Hybrid Populus (*P. deltoides* x *P. trichocarpa*) grown at three contrasting sites – Linking genetic information to the physical sequence. Pretoria, South Africa

Rae AM, Pinel MPC, Bastien C, Sabatti M, Street NR, Tucker J, Dixon C, Marron N, Dillen SY, Taylor G (2006) QTL for yield in bioenergy *Populus*: identifying GxE interactions for growth at three contrasting sites. Abstract Book p. 43. Fourth International Poplar Symposium, Nanjing, China

Rae AM, M.P.C.Pinel, C. Bastien, M. Sabatti, G. Scarascia-Mugnozza, G. Taylor (2004) Using the POPYOMICS experimental network across contrasting European climates to identify robust QTL for adaptive traits. Poster presented at the 12th New Phytologist Symposium. Gatlinburg Tennessee, USA

Rae AM, R. Ferris, S. Bunn, G.A.Tuskan & G. Taylor (2004) QTL detection for poplar response to elevated CO_2. Poster presented at the 12th New Phytologist Symposium. Gatlinburg Tennessee, USA

Rae AM, Sewell MM, Robinson KM, Taylor G. QTL (2003) discovery in biomass poplar grown in short rotation. Poster presented at Tree Biotechnology 2 Conference. Umea, Sweden

Rae AM Robinson KM Taylor G. QTL (2001) mapping in poplar for short rotation coppice. Poster presented at IUFRO Conference 'Tree Biotechnology in the next Millennium' Stevenson, Washington, USA

Rinaldi C, Kohler A, Duplessis S, Frey P & Martin F (2004) Isolation and expression of a host-response gene family encoding thaumatin-like proteins in incompatible poplar-rust fungus interaction. *12th New Phytologist Symposium on Functional Genomics of Environmental Adaptation in* Populus. Gatlinburg Tennessee, USA

Rohde, A., Ruttink, T., Storme, V., Druart, N., Bhalerao, R., and Boerjan, W. (2004) Functional genomics of bud development and dormancy in poplar. 12th New Phytologist symposium "Functional Genomics of environmental adaptation in Populus", Gatlinburg, Tennessee, USA

Sabatti M, Bastien C, Fabbrini F, Gaudet M, Jorge V, Paolucci I, Rohde A, Ricciotti L, Taylor G, Scarascia Mugnozza G (2006) Genetic dissection of phonological traits during budset in poplar. 51st Annual Congress of the Italian Society of Agricultural Genetics (SIGA). Riva del Garda (TN), Italy

Street NR, Skogström O, Jansson S, Rae, AM, Pinel MPC, Tricker, PJ, Taylor Gail (2004) Microarray, QTL and ELP: can we use them in poplar to identify adaptation to drought? 12th New Phytologist Symposium. Functional genomics of environmental adaptation in *Populus*, Gatlinburg, Tennessee, USA

Street N R, Skogström O, Sjödin A, Tucker J, Rodríguez-Acosta M, Nilsson P, Jansson S, Taylor G (2005) Integrating the genetics and genomics of drought response in *Populus*. IUFRO Tree Biotechnology, Pretoria, South Africa

Taylor G, Trewin H, Graham LE, Street NR, Rae AM, Tricker PJ, Zaina G and Morgante M (2005) Linking physical and genetic information to understand leaf development in *Populus*. IUFRO Tree Biotechnology, Pretoria, South Africa

G Taylor, R Ferris, KM Robinson & Rae AM (2002) QTL for leaf cell and stomatal traits in elevated CO_2 in Populus. Can we move to candidate genes? Poster presented at the Society of Experimental Biology Annual General Meeting

Trewin H, Graham LE, Street NR, Rae AM, Tricker PJ, Zaina G , Morgante M and Taylor G (2006) Why are some leaves Big? University of Southampton Postgraduate conference, Southampton

Trewin H, Zaina G, Morgante M, Rae AM, Graham LE, Street NR, Wagstaff C, Dixon MS, Steenackers M, Tricker PJ and Taylor G (2006) Why are some leaves big? Linking transcript, QTL and LD approaches to understand the genetic control of leaf size and shape in *POPULUS*. Fourth International Poplar Symposium, Nanjing, China

P.J. Tricker, Rae AM, L. Graham, S. Jansson, G. Taylor (2004) Genomic regions determining environmental adaptation of leaf development in Populus. Poster presented at the 12th New Phytologist Symposium. Gatlinburg Tennessee, USA

4. Masters, PhD and other theses produced from research in Popyomics

Acquah S (2005) Molecular genetic research on *Populus* trees: aligning the genetic map with the genome sequence, and identifying genes for wood formation. Master thesis in Biotechnology, University of Gent

Fabbrini F (2007) Studio fenologico della formazione della gemma apicale in pioppi autoctoni. Master thesis, University of Tuscia, Viterbo, ITALY

Gaudet M (2006). Molecular approach to dissect adaptive traits in native European Populus nigra L.: construction of a genetic linkage map based on AFLP, SSR, and SNP markers. PhD thesis, University of Tuscia, Viterbo, ITALY

Koch B (2007) Stomatal characteristics of fast growing poplars toward renewable bioenergy. M.Sc. thesis, University of Antwerpen, Department of Biology, Wilrijk, Belgium

Neckebroeck P (2006) Role of FT proteins in the seasonal growth of poplar.
Master thesis in Biotechnology, University of Gent

Pattyn P (2005) Promoter analysis of genes involved in bud dormancy of poplar. Master thesis in Biotechnology, University of Gent

Rewers T (2006) Determination of carbon and nitrogen content of different poplar leaves. Master thesis in Chemistry, option Biochemistry. College Karel De GroteHogeschool, Hoboken, Belgium

Ricciotti, L. (2007) Valutazione di caratteri quantitativi per il miglioramento della produzione di biomassa in Populus alba L. PhD thesis, University of Tuscia, Viterbo, ITALY

Rinaldi C (2003-2007) Caractérisation des déterminants géniques impliqués dans l'interaction entre un agent phytopathogène, *Melampsora larici-populina*, et le peuplier. Thèse de doctorat, Université Henri Poincaré Nancy 1, soutenue le 30 janvier 2007

Rodriguez-Acosta M (2006) The ecophysiology of drought tolerance in *Populus*. Ph.D, University of Southampton

Street NR (2005) The genetics and genomics of drought response in *Populus*. Ph.D, University of Southampton

Tallis MJ (2007) Autumnal senescence in a poplar plantation in response to elevated CO_2, from cell to canopy

Tucker J (2006) The genetics, genomics and physiology of the response of *Populus* to elevated ozone. Ph.D, University of Southampton

5. Any other dissemination activities to the wider public, meetings, events, growers

Liberloo, M. and Dillen, S.Y. (2006). Energy from poplar and willow. Article in Dutch 'Energie uit populier en wilg' in popularizing journal EOS 2, 32-37.

Trewin H, Rae AM, Street N & Taylor G. BIODIVERSITY to BIOENERGY (2005). Biomass For Energy Genetic Improvement Network Shakeholders meeting, Southampton. This stakeholder meeting was used to disseminate the research of POPYOMICS to a grower and technology audience.

Professor Taylor was an invited speaker at the John Innes conference on Bioenergy, Norwich, September 2006, presenting a talk on 'Poplar as a bioenergy tree'.

Dr Nat Street was part of a UK delegation to China on renewable energy where his expertise in Bioenergy was used to foster links with Chinese scientists and technologists.

Professor Taylor undertook several radio and TV appearances including an interview on German radio 'Making the perfect Bioenergy Tree' , May 2007and Meridian TV programme on Bioenergy crops in the UK, June 2007.

Professor Taylor was an invited speaker and contributor to the ‘Genomics and Society’ ESRC Genomics Forum, presenting a talk entitled ‘Genomics for Bioenergy’ April 2007, Cambridge.

Professor Taylor, in her role to cover Bioenergy in UKERC organised and lead the UK Bioenergy roadmapping workshop, Oxford University, The Meeting Place, April 2007, from which the UK Bioenergy roadmap will be developed.

Professor Taylor contributed evidence to the Science and Technology Select committee on Renewable Energy, as part of the UKERC submission, June 2007.

Professor Taylor was an invited keynote speaker at the Gatsby Summer School in Plant Biology speaking on ‘Plants for Bioenergy’, July 2007.

7. Policy Related Benefits

Policy development in the field of renewable energy in the EU is currently extremely active. Several of the output of POPYOMICS are relevant to the development of second generation lign-cellulose feedstocks such as poplar as a source of renewable, sustainable energy for heat, power and liquid transportation.

Bioenergy has an important role to play in meeting the EU aspirations in renewable energy supply for 2010 and 2020. Energy from biomass is complicated since several feedstocks (e.g. dedicated bioenergy crops such as willow, or food crops such as rape, sugar beet and wheat) may be utilised in different conversion processes (combustion, fermentation, gasification) resulting in several energy outputs including ***heat, power*** and ***liquid transport fuels*** (called here biofuels). It is important that this mixed portfolio of bioenergy supply is maintained at this time, ensuring the development of competitive and secure bioenergy and a firm research base for future large-scale deployments.

Deployment is being encouraged by a large number of EU incentives including for co-firing, energy crop planting grants and capital programme incentives and these may extended further. Research to develop high yielding feedstocks and improvement of inefficient conversion processes and environmental sustainability is also developing, enabling us to identify clear short-term research priorities for the EU and research in POPYOMICS is of direct relevance to that ambition.

There are powerful, long-term environmental, political and economic drivers for the further development of the international bioenergy sector. Bioenergy development in the past largely by the persistent low cost of crude oil and associated policy and development barriers. The economic situation has undoubtedly changed and at $90 per barrel, many bioenergy operations begin to approach a commercial reality. We have entered a period of sustained high fossil oil prices, with a seemingly long-term upward trend and in the future a move towards a more 'bio-based' economy where bio-based products (including bioenergy) are seen to have a higher value. Much advanced research will be necessary to make this move to biofuels, bio- polymers and bio-

oils as well as other products, over the next two decades[3], within the appropriate framework of environmental and economic sustainability.

Current emphasis on the use of food crops (first generation bioenergy crops) to generate bioethanol and biodiesel has highlighted the potential environmental costs and poor energy balance of these approaches and these must be addressed through a move towards second generation perennial lignocellulosic and other more efficient systems and a reconsideration of biomass use for heat and power compared to liquid biofuels.

There is clear strategic vision in Europe through the EU and the United States, which is being matched by considerable resource investments, not least at the biology end of the R&D spectrum to underpin these policy intiatives.

References

BERR UK Energy Statistics 2007

The Path Forward for Biofuels and Biomaterials -- Ragauskas et al. 311 (5760): 484 -- Science

EPOBIO- Bioproducts from non-food crops

An EU strategy for Biofuels

Genomes to Life: Systems Biology For Energy and Environment

Breaking the biological barriers to cellulosic ethanol. A joint research agenda. A research roadmap resulting from the biomass to biofuels workshop, Dec 2005, pblished by DOE, June 2006.

Waste strategy for England 2007

EU Biomass Action plan

Report: Biofuels in the European Union - A Vision for 2030 and Beyond on the ManagEnergy Website

Biomass Task Force, October 2005, Report to UK Government

Farrell AE, Plevin RJ, Turner BT, Jones AD, O'Hare M, Kammen DM. 2006. 311, 506-508, Science,

Rowe, R, Street NR and Taylor (2007). 'Identifying potential environmental impacts of large-scale deployment of dedicated bioenergy crops in the UK' Renewable and Sustainable Energy Reviews (in press)

UK Biomass Strategy, 2007

Bioenergy Funders Forum. Research priorities – 2007.

EU Biofuels Progress Report, 2007

How much bioenergy can Europe produce without harming the environment? - English - EEA

EBI BP Energy Bioscience Institute

DOE, USA Bioenergy Centres, 2007