Tree Physiology 34, 1167–1180
doi:10.1093/treephys/tpt089

Invited review

Genetic strategies for dissecting complex traits in biomass willows (*Salix* spp.)

Steven J. Hanley and Angela Karp[1]

Department of AgroEcology, Rothamsted Research, Cropping Carbon Institute Programme, Harpenden, Hertfordshire AL5 2JQ, UK; [1]Corresponding author (angela.karp@rothamsted.ac.uk)

Received May 20, 2013; accepted September 8, 2013; published online November 10, 2013; handling Editor Ulrika Ganeteg

Willows are highly diverse catkin-bearing trees and shrubs of the genus *Salix*. They occur in many growth forms, from tall trees to creeping alpines, and successfully occupy a wide variety of ecological niches. Shrubby willows (sub-genus *Vetrix*) have many characteristics that render them suited to cultivation in much faster growth cycles than conventional forestry. They respond well to coppicing, can be propagated vegetatively as cuttings and achieve rapid growth with low fertilizer inputs. As a result, willows grown as short rotation coppice are now among the leading commercially grown biomass crops in temperate regions. However, although willows have a long history of cultivation for traditional uses, their industrial use is relatively recent and, compared with major arable crops, they are largely undomesticated. Breeding programmes initiated to improve willow as a biomass crop achieved a doubling of yields within a period of <15 years. These advances were made by selecting for stem characteristics (height and diameter) and coppicing response (shoot number and shoot vigour), as well as resistance to pests, diseases and environmental stress, with little or no knowledge of the genetic basis of these traits. Genetics and genomics, combined with extensive phenotyping, have substantially improved our understanding of the basis of biomass traits in willow for more targeted breeding via marker-assisted selection. Here, we present the strategy we have adopted in which a genetic-based approach was used to dissect complex traits into more defined components for molecular breeding and gene discovery.

Keywords: **bioenergy, breeding, genetics, genomics, *Salix*, quantitative trait loci.**

Introduction

Willows are catkin-bearing trees and shrubs of the genus *Salix*. Together with *Populus* (poplar, aspen and cottonwood), they constitute the family Salicaceae s. str. of the order Salicales, class Magnoliopsida, subclass Dilleniidae. They are extremely diverse and show a wide variety of growth habits, from tall trees to shrubs and creeping or dwarf forms. There are about 330–500 species (depending on the classification system adopted) that are broadly grouped into the tree willows (sub-genus *Salix*), the dwarf and alpine willows (sub-genus *Chamaetia*) and the shrubby willows (sub-genus *Vetrix*) (Newsholme 1992, Argus 1997). Collectively, these species occupy a wide variety of ecological niches, mostly in temperate and arctic zones, although some willows have adapted to subtropical and tropical regions and a few are native to the Southern Hemisphere. Asia is considered to be the centre of origin of *Salix* and around 275 species are found in China, of which 189 are endemics. Around 120 species are found in the former Soviet Union, over 100 in North America (one species is native to South America) and around 65 species in Europe (Argus 1997). This wide diversity is reflected in the genetics, biochemistry, physiology and morphology of willow and is expanded yet further by the wide degree of inter-specific hybridization that occurs within the sub-genera and the plasticity that can be shown by willows in different environments.

Willows have a long history of cultivation that arises from many historical and traditional uses (Karp 2013). The most important and well known of these are: basketry, for which there are records of willow cultivation dating back to Roman times; cricket bats, in which the use of white willow dates back to the late 1800s; and the treatment of ailments, which was recognized by the Greeks in 5th century BC and culminated in the manufacture of aspirin by Bayer in 1899. From these traditional uses, basic knowledge has accrued on growth and performance and also on cultivation practises such as pollarding and coppicing, in which willow stems are planted, allowed to grow and then cut back in repeated cycles. In addition, many species were grown locally under pseudo-varietal names particularly for basket-making, and although these were mostly indigenous species, there was also export and import among countries world-wide, and inter-specific hybridization occurred. Basketry has utilized the shrubby willows, with their flexible and often coloured stems, and ease of coppicing. Species such as *Salix triandra* L., *S. purpurea* L. and *S. viminalis* L. are among the many favourites cultivated in Europe, *S. eriocephala* Michx, *S. cordata* Muhl. and *S. purpurea* L. in North America and the *S. kinuiyanagi* Kim. in Japan. In other countries, such as India, which has no native species, willows like *S. viminalis* L. were imported from the UK for basket-making and *S. alba* L. (the 'English willow') for cricket bats.

More recently, willows have become of interest as a potential source of sustainable and renewable biomass for the bioenergy, biofuel and bioproduct industries. These applications exploit some key properties of willow that are found especially in the shrubby species of the sub-genus *Vetrix*: fast, vigorous growth in coppicing cycles and ease of propagation and low fertilization requirements (an average of 20–30 kg N ha^{-1} $year^{-1}$, depending upon site). These properties render shrubby willows more suitable for cultivation on a wider scale and within time frames that are much shorter than traditional forestry. When grown in a typical short rotation coppice (SRC) cycle, short cuttings of willow stem (~15 cm tall) are planted in spring and grown for 1 year, during which time they can reach 2–3 m in height. Once the leaves have dropped, the stems are cut back to induce a coppicing response, in which multiple shoots re-sprout from the cut base (stool) in the following spring. Growth without harvest is allowed for typically three more years, by which time the stems can reach in excess of 7 m tall. The stems are harvested and chipped or billeted for storage and use at industrial plants. In the following spring re-sprouting of new shoots occurs from the cut stems (stools) and the cycle is continued for ~20 years (Volk et al. 2006, Karp and Shield 2008).

The new interest in willow as an industrial crop encouraged the initiation of breeding programmes, particularly in the UK, Sweden and North America. As with all crops, emphasis was placed on achieving yield gains through both agronomic and genetic improvements. This included much early work on defining optimized planting densities, lengths of rotation and harvest cycles (Verwijst 1996*a*, 1996*b*) as well as assessing the performance of promising species or natural hybrids, and the progeny of new genetic crosses in yield trials (Lindegaard and Barker 1997, Larsson 1998, Lindegaard et al. 2001). Over the years, genetic and genomic approaches, combined with both extensive and intensive phenotyping, have substantially improved our understanding of the basis of biomass traits in willow for more targeted breeding via marker-assisted selection (Karp et al. 2011). Here, we describe the genetics-based strategy that we have adopted to dissect complex traits into more defined components for molecular breeding and gene discovery.

Initial target traits

When willow first became recognized as a promising biomass crop, Stott (1984) characterized 54 species from the UK National Willow Collection (Figure 1a) for basic growth traits and identified shrubby willows, particularly *S. viminalis* as the most suitable. Similarly, in Sweden, Larsson (1998) identified several shrubby species as promising biomass crops, including *S. viminalis* L. but also *S. dasyclados* Wimm., *S. schwerinii* E. Wolf., *S. triandra* L., *S. caprea* L., *S. daphnoides* Vill. and *S. eriocephala* Michx. The primary target traits identified at this time were biomass yield and resistance to rust. Yield was assessed non-destructively as stem height and diameter (or stem cross-sectional area) and destructively as biomass harvest, expressed yearly as oven-dried tonnes per hectare per year (odt ha^{-1} $year^{-1}$). Rust, caused by *Melampsora* spp., is the major pathogen of willow and exists as several pathotypes of both leaf-infecting and stem-infecting forms (Pei et al. 1996, Ramstedt 1999*a*, 1999*b*, 2002). Most early selections were based on field resistance, with little knowledge of the pathotypes present. Secondary targets for crop improvement included coppicing response (response to cutback, number of shoots and shoot vigour), stem straightness (which is important for machine harvesting) and resistance to pests (such as willow beetles), other diseases (such as stem die back) and environmental stresses (cold, drought) (Larsson 1997, 1998, 2001). Due to efforts based on selection for all these traits, biomass yields of willow doubled from <7 odt ha^{-1} $year^{-1}$ to exceed 14 odt ha^{-1} $year^{-1}$ in ~15 years. These increases were mostly achieved by selection on the direct hybrids of intra- and inter-specific crosses, although recurrent selection strategies were explored (Gullberg 1993). They were also achieved mostly on the basis of gross phenotype, with limited understanding of the genetics or physiology of any of the key traits.

Some attention also turned to assessment of compositional traits, as willow biomass became utilized as a feedstock for heat and power stations. The calorific value of willow wood chips is 19.7 MJ kg^{-1} (Ledin 1998), which is similar to wood

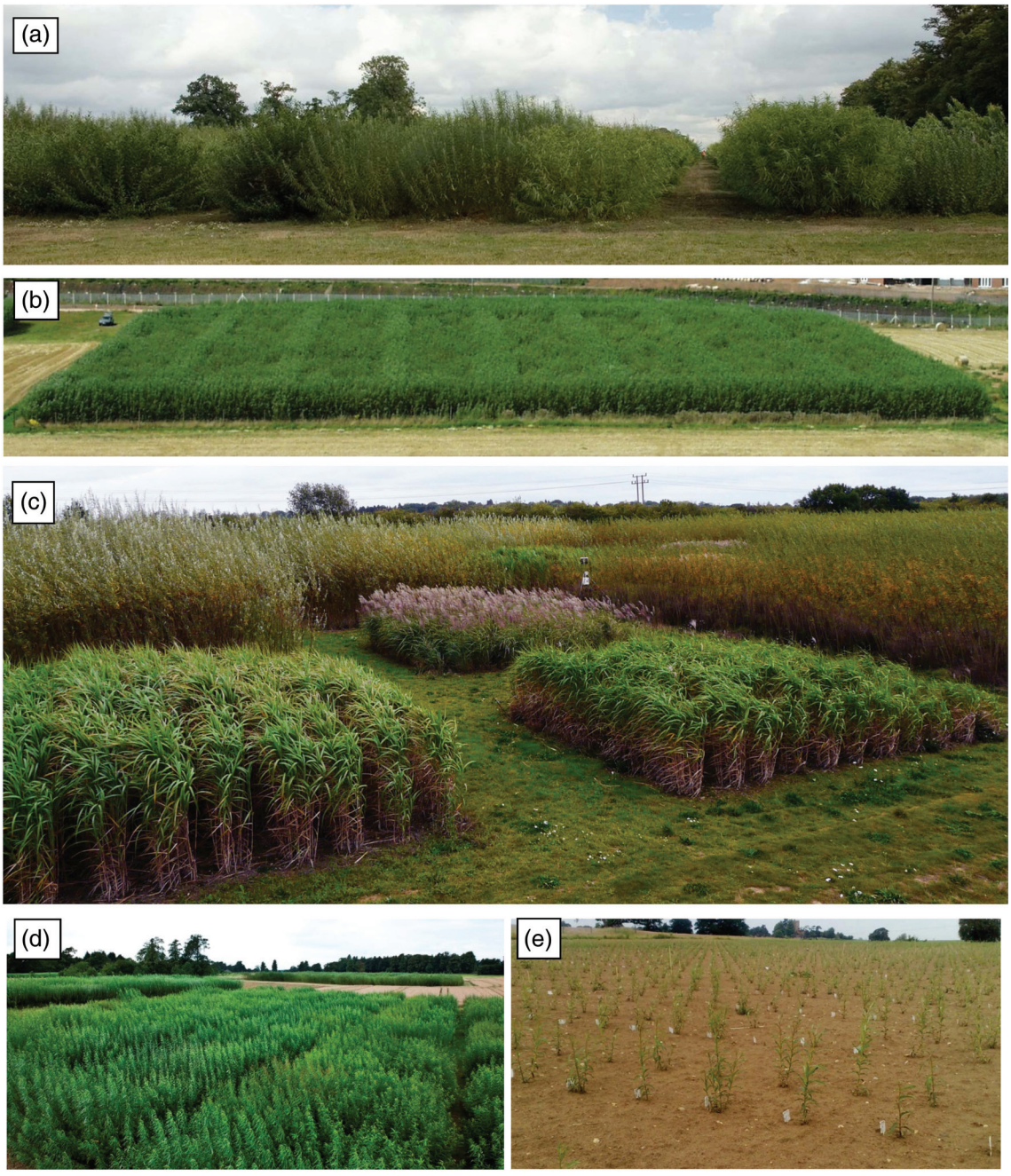

Figure 1. Genetic resources developed in willow. (a) Partial view of the National Willow Collection (NWC) at Rothamsted Research (RRes). The NWC was initiated in 1922 and now contains ~1400 accessions, including 100 pure species. It is maintained as a coppiced collection. (b) The K1 and K8 mapping populations at Long Ashton Research Station (LARS); adjacent in a single block. This trial was planted in 2000 and replicates of K8 only were subsequently planted at RRes in 2002 and at the Rothamsted Woburn farm in 2009. (c) The dedicated trial planted for the BBSRC-funded BSBEC project for destructive and non-destructive measurements, comprising four contrasting genotypes of willow and of *Miscanthus* and planted at RRes (south-east) and IBERS (south-west). (d) Partial view of the 11 additional mapping populations at RRes. (e) Partial view of the association mapping population after establishment at the Woburn site (well-drained nutrient-poor soils).

from conventional forests (19.4–21.2 MJ kg^{-1}). The ash, sulphur and phosphorus contents are low (1.0–1.2, 0.03 and 0.09%, respectively) compared with straw, for example, which is advantageous as these elements cause problems in the combustion process. Data were also collected on percentage lignin, cellulose and hemicellulose, typically 25, 30 and 40%, respectively, although genotypes vary (Ray et al. 2012, Serapiglia et al. 2013, Stolarski et al. 2013). These traits have not yet been subjected to intensive selection programmes although they have become targets of greater interest, as willow has been considered among the potential lignocellulosic feedstocks for biofuels as well as bioenergy.

Most of the traits relating to both biomass yield and composition, as well as to the plants' ability to withstand stress, are complex traits, for which many genes may be contributing often minor effects and for which the environment may play a key role. This makes the identification of key genes and the practise of selection more difficult to achieve. For further improvement of willows, such complex traits needed to be broken down into subcomponents for which environmental influences may be lower and the heritability higher and which can therefore be mapped more precisely in genetic analyses for marker and gene discovery purposes. Here, we demonstrate that through the application of a genetics strategy it is possible to uncover more about the biology of the traits themselves as well as provide the tools for marker-assisted selection.

Challenges and opportunities for genetic approaches in willow

There has been a technological revolution in the way that genomes and their expression can be interrogated and genomic regions affecting traits can be searched for. Three main routes are often used in combination: (i) characterization of transgenic lines or variants identified by screening germplasm, natural or mutant populations and ecotypes; (ii) genetic mapping to identify quantitative trait loci (QTL) and genes through linkage or association genetics; and (iii) comparison of transcripts, metabolites or proteins from different tissues, developmental stages and/or genotypes.

The huge diversity present in willow constitutes a gold mine for efforts aimed at genetic improvement, but there are significant challenges to be faced in deploying any of the above strategies. Although some species of interest are diploid ($2n = 2x = 38$), many New World and European species (including many biomass varieties) are polyploid, with ploidy levels that can reach up to dodecaploid ($2n = 12x = 228$) (MacAlpine et al. 2008). In addition, although most species within the sub-genera hybridize readily, crossing barriers do exist (MacAlpine et al. 2008) and the dioecy can be a problem if it transpires that genotypes desired for crossing are of the same sex. Few traits can be scored with any reliability in the nursery and although non-destructive methods can be used to estimate yield, true (harvested) biomass yield can only be assessed after 2–3 years and then in successive SRC cycles (Figure 2). Measuring in a mature coppice plantation is time consuming and challenging, and trials are expensive to maintain on multiple sites and to assess for long periods of SRC cycles. Since willows are dioecious and cannot be selfed to form inbred lines, genetic and genomic approaches, and the subsequent analyses of their outputs, have to deal with high levels heterozygosity. Moreover, it is known from comparison with poplar (Hanley et al. 2006, Berlin et al. 2010) that *Salix* has undergone the same recent genome duplication event, thus, being certain that sequences originate from the gene of interest and not a paralogous locus can be problematic. Finally, there is currently no robust method for transformation and thus no means of validating gene function via the production of transgenic plants of willow itself.

Given such challenges, genetic and genomic approaches need to be smart and to exploit the advantages that willow does offer over many other tree species. Unlike poplar, for example, where it may take several years for plants to flower, most willows will flower within the first year of growth from a cutting or seed and it is usually easy to create large numbers of progeny. Propagation via cuttings means that the same genotypes can be replicated on different trial sites. Genetic analysis has shown that recombination rates are high in willow and linkage disequilibrium can be fairly low (Berlin et al. 2011), facilitating mapping to high resolution. Although, at present, the large number of markers that would be required for a comprehensive genome-wide screen has yet to be generated in willow, and suitably high-throughput and cost-effective screening methods are not yet available, a candidate gene approach is feasible and advances in sequencing technologies may facilitate the former in coming years (see later). With these points in mind we have pursued the following genetic strategy to unravel complex traits in willow.

Genetics and genomics strategy

Creation of an initial map anchored to the poplar genome

Although the first genetic maps for willow were published in the early 2000s (Tsarouhas et al. 2001, Hanley et al. 2002, Rönnberg-Wastljung et al. 2003), the limited progeny sizes of the crosses used were considered too small for informative trait mapping. To increase the chances of mapping QTL with a high-enough resolution that would facilitate the identification of candidate genes, we recognized that it would be important to have a sufficiently large population size. In the late 1990s, there was little prior information on trait segregation in willow; therefore, to generate an informative population with increased certainty, two crosses were performed between parents carefully selected on the basis of both pedigree and phenotypic traits, and two families (K1 and K8; Table 1) of up to 1000 progeny reared from each and planted in the year 2000 in a field trial at Long Ashton Research Station (LARS: 51°25′22″N, 2°40′12″W; 50 m above ordnance datum (AOD)) in the UK. Both K1 and K8 families were assessed for trait variation for a full SRC cycle, and based on the patterns of segregation for the main traits of interest K8 was selected for genetic mapping (Figure 1b). After some losses, K8 eventually comprised 947 progeny of a cross between two complex hybrid full-sib *S. viminalis* × *S.schwerinii* individuals. The K8 population is still maintained at LARS, but replicates of the whole trial were subsequently planted at Rothamsted Research (RRes: 51°48′30″N, 0°21′22″W; 125 m AOD) in 2002 and at the

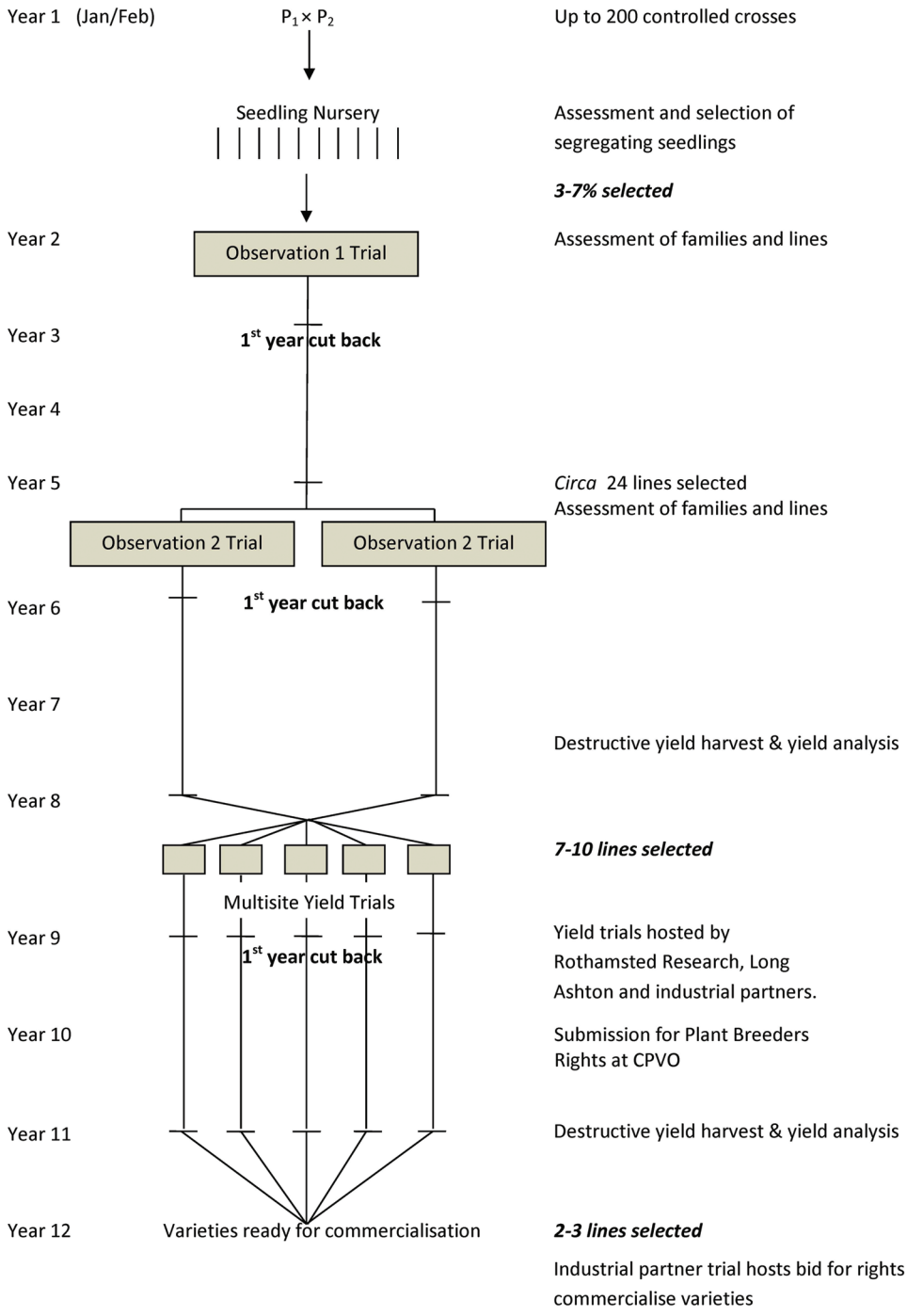

Figure 2. Schematic representation of the willow breeding programme at Rothamsted Research, from initial crossing to multi-site trialling.

Rothamsted farm at Woburn (52°0′43″N, 0°35′36″W; 95 m AOD), which has well-drained, nutrient-poor soils, in 2009. These trials, on contrasting soils and in contrasting east–west climatic conditions, have been the focus of detailed trait assessments over successive coppicing cycles.

Using 467 of the K8 progeny, a detailed linkage map was constructed using SSR, SNP and AFLP markers. To facilitate transfer of information from poplar the K8 map was directly aligned to the poplar genome (Hanley et al. 2006). To achieve this, we selected a set of genome-wide, expressed poplar sequences and used these to design primer sets that efficiently amplified homeologous regions in willow. Single-nucleotide polymorphisms (SNPs) were used to map the loci and align the willow linkage groups to the poplar genome sequence. This was the first report of the high degree of macrosynteny between willow and poplar (Hanley et al. 2006). At the time of publication, coverage was 1856.7 cM with an average interval of 6.3 cM, but this has since been improved significantly in

Table 1. Bi-parental *Salix* mapping populations established at Rothamsted Research.

Population	Pedigree	Created	Field trial location (date established)	*n*	Map	Markers
K3[a]	*S. viminalis* L. (SW870084) × *S. viminalis* L. (SW87002) 'Orm'	1994/99	LARS (2000)	208	Y[b]	SSR, AFLP
K1	Complex *S. viminalis* L. × *S. schwerinii* E. Wolf hybrid siblings 'S1 × R13'	1999	LARS (2000)	946	N	–
K8	Complex *S. viminalis* L. × *S. schwerinii* E. Wolf hybrid siblings 'S3 × R13'	1999	LARS (2000), RResH (2002), RResW (2009)	947	Y[c]	SNP, SSR AFLP
mpA	K8-411 × *S. viminalis* L. (NWC674)	2007	RResH (2008)	480	Y	SNP, SSR
mpB	*S. viminalis* L. (NWC663) × K8-165	2007	RResH (2008)	480	N	–
mpC	*S. viminalis* L. (NWC655) × K8-165	2007	RResH (2008)	480	N	–
mpD	*S. viminalis* L. × *S. purpurea* L.(NWC789) × K8-165	2007	RResH (2008)	548	Y	SNP, SSR
mpE	*S. caprea* L. × *S. cinerea* L. × *S. viminalis* L. (NWC901) × K8-165	2008	RResH (2009)	626	Y	SNP, SSR
mpF	*S. x alberti* L. (*S. integra* Thunb. × *S. suchownesis* W.C. Cheng ex G.Zhu) × K8-290	2008	RResH (2009)	394	Y	SNP, SSR
mpG	K8-319 x (*S. viminalis* × *S. repens*[d])(NWC1059)	2008	RResH (2009)	593	Y	SNP, SSR
mpH	(*S. viminalis* L. × *S. schwerinii* E.Wolf) 'Tora' × *S. daphnoides* Vill. (NWC432).	2008	RResH (2009)	494	N	–
mpI	*S. aurita* L.(NWC453) × K8-290	2008	RResH (2009)	407	N	–
mpJ	*S. schwerinii* E.Wolf. (NWC615) × (*S. viminalis* L. × *S. schwerinii* E. Wolf) 'Bjorn'	2008	RResH (2009)	261	N	–
mpK	(*S. viminalis* L. × *S. schwerinii* E. Wolf) 'Tora' × *S. triandra* L (NWC099).	2008	RResH (2009)	181	IP	SNP, SSR

K8-xxx, member of K8 family used as parent in later crosses; LARS, Long Ashton Research Station, Somerset, South West UK; RResH, Rothamsted Research, Harpenden, South East UK; RResW, Rothamsted Research, Woburn Farm, South East UK; Nutrient poor, drought prone site, ± irrigation; IP, in production.

[a]Original cross by Svalöf Weibull AB (Sweden) but cross later re-made at LARS to increase progeny size to 209 from 66.

[b]Full details available in Hanley et al. (2002).

[c]Full details available in Hanley et al. (2006).

[d]National Willow Collection Records incomplete—authority unknown and identification under enquiry.

targeted regions of interest. A subsequent study that employed additional genome-wide connecting markers has added to our knowledge in this area (Berlin et al. 2010).

Using K8, QTLs have already been mapped for a number of traits, including biomass yield and some of its components (i.e., shoot height, diameter and number), rust resistance (Hanley et al. 2011) (Table 2), and more recently phenology, water-use efficiency, drought tolerance, composition (S. J. Hanley, unpublished results, papers in preparation) and saccharification potential (Brereton et al. 2010). A range of QTLs in willow have also been mapped in this way by other research groups, although population sizes have been smaller than K8, (e.g., Barcaccia et al. 2003, Tsarouhas et al. 2003, Tsarouhas et al. 2004, Rönnberg-Wastljung et al. 2005, 2006, 2008, Lin et al. 2007, Liu et al. 2011, Samils et al. 2011).

Recognizing the importance of phenotypic data from long-term trials

In an era of 'omic' tools and technologies, it can be easy to overlook the importance of robust phenotypic data, but the value of this could never be over-emphasized. Although they take many years to accrue in perennial biomass crops and are extremely resource intensive to generate (for example, a full biomass yield assessment of K8 can take six people working for 4–6 weeks) our phenotype data sets, collected as comparable measurements over multiple years and harvest rotations, and from different, contrasting field sites, remain one of our most important assets. They enable us to continue to search for new QTL once target traits of interest are identified, or better refined, and to explore temporal stability as well as gene × environment (G×E) interactions. In most cases, multiple QTL are revealed for the same trait and information on the consistency of their contribution to explain the variation in the trait over time and environment can help identify which ones to focus on as priority targets for further study. Conversely, site-specific loci may prove particularly informative if they can be linked to particular site characteristics.

Mapping multiple traits in the same population can also provide valuable insight into the contribution of different individual QTL to the more complex overarching trait. For example, the QTL data presented in Table 2 have allowed us to further dissect biomass yield in terms of understanding the contribution of particular genomic regions to various yield components, through the co-location of QTL for different, but correlated traits. For example, a QTL for biomass yield co-locates with the major K8 rust resistance locus *SRR1* (Hanley et al. 2011), on linkage group Ib. Similarly, a biomass yield QTL on linkage group X influences stem height and stem diameter but does not exert a detectable effect on shoot number. Such results can help focus the selection of candidate genes in situations where the simpler component trait is better understood at a fundamental level, with specific genes or gene classes already implicated in the relevant processes.

It can be difficult to carry out all the phenotypic trait analyses needed on mapping populations, as measurements of

Table 2. Summary table of QTL identified for biomass yield and component traits in the K8 willow mapping population at two different sites (LARS, Somerset, UK and RRes, Hertfordshire, UK) over two successive harvest cycles.

	First cycle (2 years: 2001–2002)							First cycle (2 years: 2003–2004)				Second cycle (3 years: 2003–2005)				Second cycle (3 years: 2005–2007)			
Site	LARS							RRes				LARS				RRes			
LG	FW'02	MxHt'02	MxDia'02	Shts'02	Rust'00	Rust'01	Rust'02	FW'04	MxHt'04	MxDia'04	Shts'04	FW'05	MxHt'05	MxDia'05	Shts'05	FW'07	MxHt'07	MxDia'07	Shts'07
Ia	–	–	–	[a]4.4	[a]3.4	[a]5.4	[a]4.7	–	3.6	–	–	[a]3.0	–	3.2	3.6	–	–	–	–
Ib	–	–	3.08	–	[b]76.4	[b]41.5	[b]48.4	–	–	–	–	[b]3.7	[b]3.1	[b]4.4	–	–	–	–	–
II	–	–	–	–	–	–	–	–	–	–	–	–	3.0	–	–	–	–	–	–
III	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	6.0	–	–	–
Va	–	–	–	–	[c]2.6	[c]3.8	[c]3.1	–	–	–	–	[d]3.3	–	–	[d]3.6	–	–	–	–
VIa	[e]4.2	[e]4.2	[e]4.1	[f]4.71	–	–	–	[e]2.5	–	[e]1.8	[e]1.9	[e]3.0	[e]3.2	[e]2.7	[f]6.4	–	–	–	–
VIc	[g]1.6	[g]1.7	[g]2.7	–	–	–	–	–	[g]1.3	[g]2.5	–	–	[g]1.5	[g]1.7	–	–	–	–	–
VIIIa	–	–	–	–	–	–	–	–	–	–	–	–	–	–	2.6	–	[h]4.3	[h]3.3	3.4
VIIIb	–	–	–	–	–	–	–	–	–	–	–	1.2	–	–	–	–	–	–	–
IX	–	–	–	[i]3.0	3.1	–	–	–	–	–	–	–	–	–	[i]3.0	–	–	–	–
X	[j]12.9	[j]23.8	[j]20.4	3.7/[i]4.7	–	–	–	[j]6.4	[j]10.5	[j]13.8	–	[j]6.1	[j]11.7	[j]10.3	[j]5.0	[j]4.0	[j]14	[j]11.1	–
XI	–	–	–	–	[k]9.8	[k]6.5	[k]6.3	–	–	–	–	–	–	–	–	–	–	–	–
XII	–	–	–	–	–	–	–	[l]2.4	–	–	[l]1.9	–	–	–	–	–	–	1.5	–
XIIIa	–	–	–	–	–	–	–	–	–	–	–	–	2.3	–	–	–	–	–	–
XIIIb	–	–	–	–	–	–	–	[m]1.5	–	[m]3.2	[m]1.3	–	–	–	–	–	–	–	–
XIV	–	–	–	–	[n]3.5	[n]3.0	–	[n]4.4	–	–	–	[n]4.0	[n]4.1	[n]3.9	–	–	4.2	[o]3.1	4.8
XVI	–	–	–	–	–	–	–	–	–	–	3.0	–	–	–	–	–	–	–	–
XVII	–	[p]2.6	–	–	–	–	–	–	–	–	–	–	–	–	–	–	[p]3.1	[p]3.7	–
XVIII	–	[q]2.1	[q]2.6	–	–	–	3.9	–	–	–	–	[q]2.7	–	[q]2.7	–	–	–	–	–
XIXb	–	–	–	–	–	2.2	–	–	–	–	–	–	–	–	–	–	–	1.7	–

Maximum LOD (logarithm (base 10) of odds) scores calculated by Interval Mapping within the MapQTL 4.0 software package are presented if significant by permutation test at the chromosome level. LOD scores also significant at the genome level are underlined. Equivalent superscript characters are indicative of putative co-locating QTL. Linkage groups and map data are as presented in Hanley et al. (2006). Linkage groups with no QTLs are not shown. Field trial details and protocols for phenotypic assessments are available in Hanley (2003).
FW, fresh weight at harvest; MxHt, maximum stem height; MxDia, maximum stem diameter; Shts, number of shoots per stool.

some traits require complete destructive sampling and removal of individual genotypes can compromise measurements of the surrounding plants. To overcome such issues and provide plants for intensive phenotyping, a dedicated trial of both willows and *Miscanthus* (*Miscanthus* spp.) was established at RRes (south-east UK) and the Institute of Biological, Environmental and Rural Sciences (IBERS), University of Aberystwyth (south-west) in Spring 2009, as part of the BBSRC Sustainable Bioenergy Centre programme on perennial biomass crops (BSBEC). Four contrasting genotypes of willow (Endurance, Resolution, Tora and Terra Nova) and of *Miscanthus* (EMI-11 and Goliath (*M. sinensis*), Sac-5 (*M. sacchariflorus)* Giganteus (*M x giganteus*)) were carefully selected to provide contrasts in growth strategies (Figure 1c). The trial was planted using a randomized block design with areas in each plot reserved for non-destructive measurements as well as for destructive measurements, where whole plants can be dug up frequently over the full lifetime of the trial. Weather stations were installed at both sites. The trials have been subjected to intensive assessment of growth and composition. More than 2 years of these BSBEC phenotypic data have been used to parameterize and validate process-based models. Together with sensitivity analysis the models have both supported and refined our original hypotheses, and helped identify the most important parameters affecting yield for crop improvement (Cerasuolo et al. 2013). Outputs from the models have been important in selecting QTL for further study, for example rate of leaf and stem extension, and above–below ground allocation rather than biomass yield per se (Karp et al. 2013).

In recognition of the importance of gathering phenotypic data, high throughput phenotyping technologies have been developed for application in laboratory and field-based platforms. It is not easy, at present, to see how the full set of available technologies could be applied to mature willow plots in the field but there are procedures, for example, based on remote sensing technologies that could be explored to facilitate data collection from large numbers of plants in different environments (Vos et al. 2010).

Creation of multiple mapping populations

The K8 population has proven extremely useful for mapping a large number of QTLs and for studying G×E interactions. However, it is only one genetic population with a restricted number of alleles segregating at any one locus. Thus, while continuing to work on K8, which remains our most developed willow map, we have expanded our QTL discovery pipeline to include new and more diverse populations (Table 1). Eleven more populations of up to 626 progeny were developed from diverse parentage and have been established in field trials (Figure 1d) and subjected to intensive phenotyping for traits of interest. Initial framework maps were constructed for these using a subset of 192 progeny and genome anchored SSRs and QTL identified. Once QTL are detected, the regions are saturated with targeted gene-derived SNP markers and mapped to maximum resolution using all the progeny. This approach has led to the generation of novel data on genome diversity in *Salix* and also enabled us to make informed choices about which QTL to target. QTL that are found across all populations are of particular interest, although population-specific QTL can also be important in capturing variation that may not have been segregating in the other families. Particular QTL that are now being pursued further from this work include several influencing timing of bud flush and senescence (S. J. Hanley, unpublished results).

An advantage of the linkage mapping approach, used for K8 and the additional populations, is that the pedigree is known but a disadvantage is that the number of recombination events are limited and closing map distances between markers delineating QTL can be difficult. Given the ease with which large numbers of seeds can be produced in willow, one approach to overcome these limitations is to remake mapping population crosses and generate very large numbers of progeny, from which seedlings can be selected using markers to identify recombinant genotypes for the target region.

Another approach is association mapping, which utilizes plants from natural populations and exploits the much larger number of recombination events that have occurred historically, enabling mapping to be achieved at much finer resolution. Spurious associations can arise due to complications with population structure, but methods exist to test and correct for this. Rothamsted collaborated with Swedish groups to form an association mapping population ($n = 369$) comprised of *S. viminalis* L. accessions from major European willow collections and new material from natural populations of the Czech Republic, Sweden and Poland. The population was planted at four sites (two Swedish, two UK; Figure 1e) and has been characterized in terms of diversity and population structure. The multi-site nature of the project generated valuable new information on G×E interactions, phenotypic stability and a resource for future association studies. Variation in several candidate genes, including loci associated with phenotypic diversity in phenology and architecture traits, is currently being analysed to identify markers for use in breeding programmes.

Development of a willow genome sequence and 'omics' approaches

Genetic and genomic approaches in willow have benefited from advances in poplar, which has been developed as both a bioenergy crop and a model tree (Yang et al. 2009). In particular, the close relationship of willow to poplar has meant that the derivation of the whole genome sequence of *Populus trichocarpa* Torr. & Gray. (Tuskan et al. 2006), and the development of many 'omic' tools and data sets, has been highly beneficial for marker development and candidate gene identification.

Nevertheless, a full genome sequence of willow itself would speed up gene discovery pipelines.

We have chosen to develop a reference sequence for willow and then re-sequence a set of key genotypes. For the reference sequence, *S. viminalis* was our choice as taxonomically it is relatively close to most of the other willow species used for biomass in the UK and has been widely used in European biomass breeding programmes. It is also the best studied species in Europe in terms of mapping key traits by conventional biparental QTL mapping and is the species that comprises the association mapping population. Sequencing of this species has now been initiated in a collaborative effort between RRes and TGAC (The Genome Analysis Centre in the UK) using a *S. viminalis* L. parent of one of the large willow mapping populations. In addition, to promote the transition from the reference genome sequence to the identification of key underlying genes, the genomes of 32 additional willow genotypes and hybrids will be re-sequenced. These genotypes include parents of the 11 mapping populations and members of a diversity panel of the *S. viminalis* L. association mapping population. This will provide the information required to characterize the genetic variants in these populations and generate a genome-wide catalogue of polymorphisms.

Transcriptomics, metabolomics and/or proteomics have been used in many plant species, including poplar, to develop a systems-based understanding of developmental changes and to compare different tissues or genotypes (Andersson et al. 2004, Bao et al. 2009, Cohen et al. 2010, Karlberg et al. 2010). The methods are resource intensive and do not normally lead to the causal genes determining a trait when used on their own, but they are very informative in showing which different sets of genes are up- or down-regulated in association with a developmental change or genetic difference. Transcriptomic studies can be achieved using microarrays in which RNA is hybridized to large numbers of genomic DNA probes (Olarte et al. 2013) or expressed sequences tags (Beritognolo et al. 2011) spotted onto solid surfaces. Rapid advances in sequencing technology have led to approaches such as RNA-seq, which has the advantages that it provides estimates of the numbers of genes and quantitative data on the abundance of transcripts. It could also lead also to the identification of novel transcripts, as the analysis is not restricted to the genes or sequences that are represented on the arrays (Marques et al. 2013, Wenger and Galliot 2013). We are now conducting transcriptomic and metabolomic studies in willow and are using these in combination with QTL analyses to identify (expression) e-QTLs (Snoek et al. 2012) or (metabolite) m-QTLs (Joosen et al. 2009).

Application of in silico mining approaches

We have simultaneously been developing software and database systems that can be used for the management and interpretation of data from genetics and genomics experiments of willow. For capturing phenotypic data from the field experiments, we have used the InterStore system (Love et al. 2012), while Ondex (Köhler et al. 2006) has been developed as an open source software framework for integrating biological data. Of particular use has been a recently developed user-friendly web interface to an Ondex application which was tuned for mining QTL and other genetic intervals for functional candidate genes using semantic data retrieval methods. This has been particularly well developed in an application called QTLNetMiner (http://ondex.rothamsted.ac.uk/QTLNetMinerPoplar), which supports the selection and prioritization of functional candidate genes under QTL intervals in the poplar genome. An example of the output generated is provided in Figure 3. Here, a QTL interval implicated in maximum plant height in the K8 mapping population at Rothamsted in 2008 (Table 2) was interrogated using 'plant height' as a keyword search and was limited to the genomic coordinates defined by sequence-based QTL flanking markers. Analysis of the current knowledge database highlighted a single gene in the interval as being linked to the search term. The network generated indicates that poplar gene POPTR_0008s00970 encodes a protein that is highly homologous to *Arabidopsis* protein PEPR1 (encoded by AT1G73080), which is known to interact with BAK1 (from AT4G33430), a rice homologue of which (OsBAK1) is known to influence architecture, including plant height, in engineered rice plants (Li et al. 2009). We have used this approach to identify candidate genes for screening in association mapping and downstream functional studies.

Validation of candidate genes

As mentioned earlier, a disadvantage of willow is that it currently lacks a robust transformation technology and there is no direct way of validating that candidates are the causal genes associated with a trait. However, the function of willow alleles can often be assessed in *Arabidopsis* or poplar transgenics.

We have not yet transformed poplar with willow genes—although if the target gene is present in poplar, RNAi could be used to knock-down the corresponding poplar orthologue. Conversely, over-expression studies may be informative in some instances. Knowledge can also be gained from transgenic poplars that have already been generated. An advantage of *Arabidopsis* is that transformation requires much less time compared with poplar and an extensive stock centre of mutant lines exists. Thus, lines that lack the gene of interest may already have been identified, which enables willow genes to be tested in a null background for the gene concerned. Additionally, a huge body of knowledge exists in *Arabidopsis* on fundamental processes associated with growth and development. In collaboration with the Leyser group in the UK, we have successfully transferred knowledge and methodologies from *Arabidopsis* to investigate the possible role of *MAX* genes in coppicing response and shoot branching in willow.

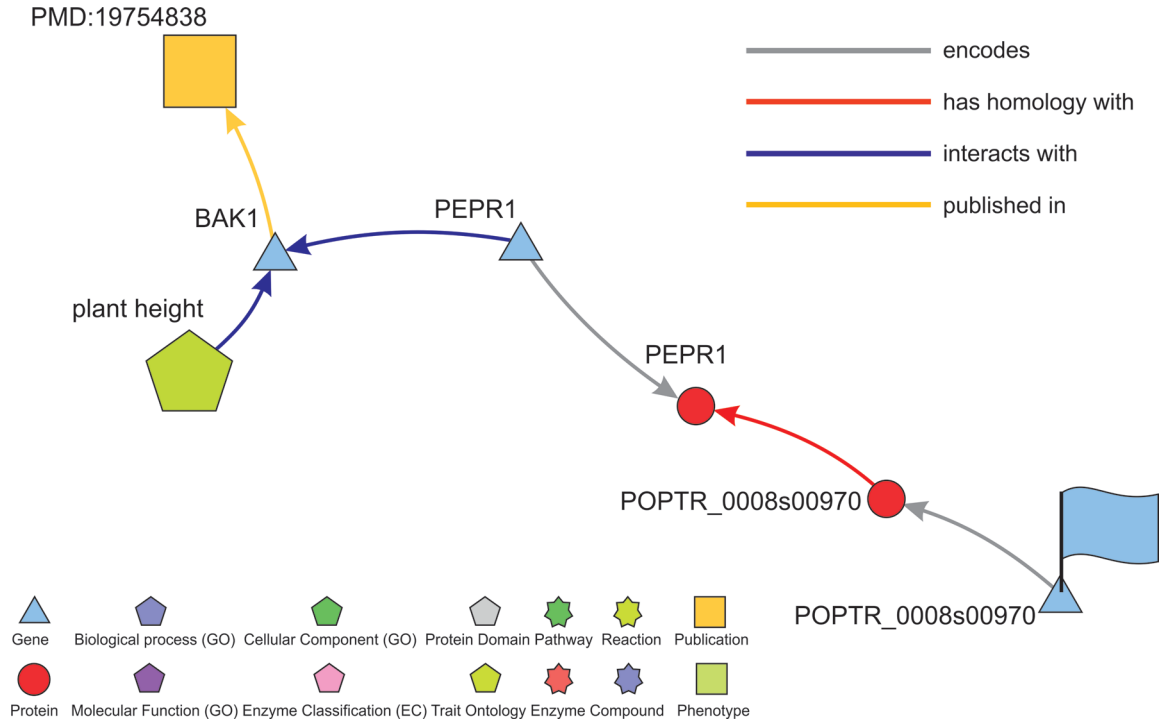

Figure 3. A QTLNetMiner-derived network highlighting poplar gene POPTR_0008s00970 as a candidate gene within a willow plant height QTL. For further explanation, see text.

Coppicing response is fundamental to the success of SRC as a biomass production system. Although variation in the extent of sprouting from cut stools is well known (Sennerby-Forsse and Zsuffa 1995), the genetic regulation of this response has not been identified. The *More AXillary growth* (*MAX*) genes are conserved among most plant species and were of interest to us in this context due to their potential role in branching. Genetic mapping revealed some association of *MAX* loci with shoot and yield QTL in willow, which is currently the study of further investigation (Table 2; positions denoted as yield QTL on linkage groups VIc and XIV). Regrowth after stem cutback in willow occurs in response to the removal of apical dominance, which has extensively been studied in *Arabidopsis*. Assays developed to study the control of bud activation and apical dominance in *Arabidopsis* were transferred to willow axillary meristems in order to assess the degree of similarity of willow responses with the known hormonal regulatory system in *Arabidopsis* (Ward et al. 2013). Despite the difference between these two plants, bud hormone responses were found to be very similar. The *Arabidopsis* bio-assay was then used for testing the function of different *MAX* alleles cloned from willow in *Arabidopsis* null lines lacking the corresponding *MAX* genes. The approach was highly successful and revealed allelic differences in willow strigolactone genes (Ward et al. 2013). We are now testing other candidates in this way.

Using genetics to refine the key target traits

Of the main routes, referred to earlier, that could be used to identify genomic regions influencing traits, our experience of the most effective strategy in willow has, until now, been to integrate knowledge from phenotyping of genetic mapping populations in field trials to identify key component traits that contribute to useful variation, and the developmental stages that are most critical to them, and then plug in the appropriate 'omic' technologies to identify the regulatory genes and markers for breeding (Karp et al. 2013). This is an iterative process in which QTL mapping can be a helpful tool in itself, enabling the co-location of more refined components traits to be investigated (see Table 2 where co-locating QTL are shown using equivalent superscript letters) and/or the mapping of new traits identified by more in depth physiological studies relative to known QTL. In addition, process-based crop models can increase the selection of improved phenotypes by identifying key component traits on the basis of their relative importance within the system.

A refined understanding of the key component trait to focus on can be critical to achieving successful outcomes, particularly in unravelling the genetic regulators of complex traits that are influenced by many variables. Examples have already been referred to earlier in discussing the advantages of mapping multiple traits in the same population, and across populations, as a means of dissecting complex traits into more defined components that have higher heritabilities for molecular breeding and gene discovery.

Canopy architecture is another complex trait in willow, which is highly variable. Willows show variations in sub-component traits such as leaf shape, leaf area, specific leaf area and leaf area index (LAI). Willow varieties selected for high biomass yield can have

quite different canopy architecture, suggesting that willows are able to achieve high yield through different growth strategies. To date results have not revealed any correlation between leaf area and yield. Process-based modelling has confirmed that the key parameters are the total leaf area (or total green area) of the plant and what has been referred to as the 'clumping index' (Cerasuolo et al. 2013). Efficient light penetration through the canopy can be achieved for willows with large LAI, if leaves also show a high degree of clumping. These results provide guidelines with which to assess canopy structures associated with different stem numbers and refined targets for breeding (Cerasuolo et al. 2013).

Another complex trait in willows of relevance to biofuels and industrial products is sugar release from the biomass. In order to identify the genes affecting this trait saccharification potential was assessed in a subset of K8 and several QTL were mapped (Brereton et al. 2010). Subsequently, QTL were mapped, also in K8, for a range of composition traits, including percentage lignin, xylan, glucan and mannan, etc. (N. J. B. Brereton, unpublished results). However, this variation did not correlate in any meaningful way with a variation in saccharification potential, nor did a variation in stem biomass or growth traits, despite the extensive number of total traits that were studied. Reaction wood (RW) was then investigated as a possible factor. Reaction wood is a fundamental response of trees to environmental and gravitational stimuli. In angiosperm trees, it is characterized by tension wood (TW), formed on the upper side of the stem, and opposite wood (OW) on the lower side. In willows, as in many other species, TW fibres may develop an inner gelatinous layer (G-fibres) that is cellulose-rich. The molecular pathways that control RW formation are little understood (Felton and Sundberg 2013), but it has been demonstrated that sugars can be more easily extracted from isolated TW compared with normal or OW (Munoz et al. 2011). As part of the BSBEC programme, we also established a beneficial saccharification effect associated with willow TW (Brereton et al. 2011). We then demonstrated that glucose release differed among the RW-induced plants, and this variation was found to be strongly correlated with the glucose release of mature field-grown trees. No such correlation was found for sugar release from the controls in this study, in which RW was not induced. The results suggest that genotypic differences in RW response may be a primary determinant of the variation observed in sugar release from willow biomass (Brereton et al. 2012). RNA-Seq is now being used to generate global transcriptome profiles for TW, OW and normal wood in two willow genotypes that differ in their response to RW induction to identify the genetic basis of the differences in response.

Future perspectives

Our Institute has been involved in willow research since 1922, originally at the LARS site and then later at RRes. The genetic improvement programme began in earnest in 1990, at a time when RFLPs and RAPDs were popular marker technologies, followed by AFLPs and SSRs and then SNPs, to mention but a few of the many marker technologies available at the time (Karp et al. 1997, 2013). Over the years, different methods of detecting DNA polymorphisms with increasing genome coverage, and of investigating the expression of genes in more global ways, have enabled new options for genetic analyses to be exploited. Using these advances, in a combination of genetic-based approaches, significant progress has been made in dissecting complex traits and in identifying genes that play key roles.

The approaches we have used and have described here for willow have a strong basis in the use of direct genetic approaches, based on both linkage and association genetic analyses, because reverse genetics has not been available in *Salix*. In this respect, the strategies have been different from those used in many other trees, where more emphasis has been given to exploiting the benefits of transformation. It could be argued that as a result of transformation being available the development of equivalent extensive genetic resources (such as the large K8 willow mapping population, established in 2002) occurred later for many trees species, including poplar. As new methods for obtaining genome-wide polymorphisms have become available, so more emphasis has been given to the collection and assembly of large populations for direct genetic approaches in a wide number of tree species particularly natural populations for genome-wide association studies (see below).

Among the biggest technological advances of recent years, now to be embraced, are those associated with genome sequencing, which not only have had massive impact on approaches for gene discovery but have also opened up opportunities to re-think strategies for genetic improvement and breeding (Harfouche et al. 2012). While existing sequencing technologies continue to evolve, and new ones emerge, the development of two, widely adopted high-throughput next-generation sequencing (NGS) pipelines by Illumina (Bentley et al. 2008) and 454 Life Sciences (Margulies et al. 2005) has perhaps been of particular significance in recent years. These latter technologies differ in output, costs and error rates, all of which have to be considered (Luo et al. 2012). Ways of using both to achieve the best of both worlds have been described (Wenger and Galliot 2013).

Next generation sequencing has revolutionized approaches to genome analysis by enabling strategies aimed at detecting all the genetic variation present to be pursued, for example, by the re-sequencing of entire genomes of specific genotypes and/or re-sequencing of specific genes or genomic regions in huge numbers of individuals (Harfouche et al. 2012). For these reasons we have now established NGS technologies at RRes.

Faster and cheaper sequencing has led to the development of many new approaches to mapping and genotyping. In genome-wide association studies, essentially large numbers of individuals are screened at very large numbers of loci for as many

phenotypes as possible, to detect associations between loci and specific phenotypes. Genome-wide association studies add considerable power to both linkage genetics and association genetics, as the highly dense genome coverage increases the chance of associations being detected and of new targets being identified for marker-assisted selection. However, some fine mapping is usually needed and validation through functional testing is still required. Moreover, even if association genetics, rather than linkage genetics is used, the approach is still limited by the strength of linkage disequilibrium between marker and locus (causing the phenotype) and by the range of alleles, and their frequencies, in the study population(s). Interesting variation present at very low frequency in the wider gene pool can be difficult to detect unless vary large populations, with rare as well as common alleles, and very large numbers of markers are used. These limitations restrict the value of these approaches to breeding where being able to capture and select for rare variation can be important (Harfouche et al. 2012).

An elegant approach in which NGS and reverse genetics were used in combination to capture rare defective alleles to accelerate breeding was described recently by Vanholme et al. (2013). A rare defective allele in lignin biosynthesis was identified by NGS in a natural population of *Populus nigra*. The mutant allele had a premature stop codon in the gene encoding hydroxycinnamoyl-CoA : shikimate hydroxycinnamoyl transferase1 (HCT1). These authors characterized the mutant phenotypes by expressing HCT1 heterologously and also made crosses to introduce the mutation in different genetic backgrounds. Trees homozygous for the recessive allele had a modified lignin composition characterized by a 17-fold increase in *p*-hydroxyphenyl units. The proposed breeding strategy, called 'Breeding with Rare Defective Alleles', should be widely applicable, independent of the target gene or the species, and could be a very powerful approach to use in willow.

Another method for accelerating breeding that has received much attention recently is genomic selection (Grattapaglia and Resende 2011). In this approach, a 'training population' is established in which all loci that regulate a phenotype are in linkage disequilibrium with at least one marker. The data from intensively phenotyping and genotyping the training population are used to develop a prediction model which is then used to predict the genomic breeding value of progeny in future generations. The genomic breeding value is essentially calculated by multiplying the number of alleles at all markers and their effect on the phenotype estimated using various regression-based statistical methods. It has been argued that based on these prediction models, genomic selection could be used to select superior genotypes early in the breeding process more effectively than marker-assisted selection and could considerably reduce the length of time required for completing a cycle of genetic improvement (Grattapaglia and Resende 2011, Harfouche et al. 2012). Success with genomic selection still has to be demonstrated in practise, but the approach has much attraction for the future breeding of willows. However, the success of using genomic selection (for all tree species) will depend on how the training population is constructed. This, in turn, requires good prior knowledge of the genetic systems (and consequent genetic variation) of the species and the populations that could be used, on the quality of the phenotypic data that are collected and on the rigour of the statistical methods deployed.

Concluding remarks

Exciting, but challenging, times lie ahead for those who wish to be successful in genetic improvement strategies beyond the exploration of the technological options themselves to the delivery phase in which improved varieties sustain their expected performance in a wide range of environments. The advancements in technologies have massively increased the power to detect polymorphisms at the DNA or transcript level and to generate vast quantities of data. Quantitative statistics and bioinformatics need to catch up with the flood of new data being generated. However, in our view, it is genetics that is still at the crux of how these advances can be most effectively utilized and the area where investment is still needed is in the creation of new long-term genetic resources that provide intensive phenotyping platforms in a wide range of environments and enable the correct trait-DNA polymorphism information to be obtained.

Acknowledgments

The authors thank William Macalpine for making the majority of willow crosses and Ian Shield, Jennifer Cunniff, Jemma Salmon, Rachel Rossiter, March Castle, Tim Barraclough, Marianna Cerasuolo, Goetz Richter, Jo Amey, Sviatlana Trybush, Stephen Powers, Keywan Hassani-Pak and Chris Rawlings for collaborative inputs on the willow research at Rothamsted Research. The authors are also grateful to Nick Brereton, Michael Ray and Richard Murphy at Imperial College, London, UK, for collaboration on willow biomass composition and reaction wood and Ottoline Leyser and Sally Ward at the Sainsbury Plant Laboratory for collaboration on MAX and branching.

Conflict of interest

None declared.

Funding

The UK Biotechnological and Biological Sciences Research Council (BBSRC) provides funding for the 'Cropping Carbon' Institute Strategic Programme Grant at Rothamsted Research; the 'BBSRC Sustainable Bioenergy Centre (BSBEC): Perennial Bioenergy Crops Programme' (BB/G016216/1: BSBEC-BioMASS) and the

Crop Science Initiative Grant (BB/E006833/1 and BB/E007007/1). Rothamsted Research is an Institute supported by the BBSRC.

References

Andersson A, Keskitalo J, Sjodin A et al. (2004) A transcriptional timetable of autumn senescence. Genome Biol 5:R24.

Argus GW (1997) Infrageneric classification of *Salix* (*Salicaceae*) in the New World. Syst Bot Monogr 52:1–121.

Bao YH, Dharmawardhana P, Mockler TC, Strauss SH (2009) Genome scale transcriptome analysis of shoot organogenesis in *Populus*. BMC Plant Biol 9:132.

Barcaccia G, Meneghetti S, Albertini E, Triest L, Lucchin M (2003) Linkage mapping in tetraploid willows: segregation of molecular markers and estimation of linkage phases support an allotetraploid structure for *Salix alba* × *Salix fragilis* interspecific hybrids. Heredity 90:169–180.

Bentley DR, Balasubramanian S, Swerdlow HP et al. (2008) Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456:53–59.

Beritognolo I, Harfouche A, Brilli F et al. (2011) Comparative study of transcriptional and physiological responses to salinity stress in two contrasting *Populus alba* L. genotypes. Tree Physiol 31:1335–1355.

Berlin S, Lagercrantz U, von Arnold S, Ost T, Ronnberg-Wastljung AC (2010) High-density linkage mapping and evolution of paralogs and orthologs in *Salix* and *Populus*. BMC Genomics 11:129.

Berlin S, Fogelqvist J, Lascoux M, Lagercrantz U, Ronnberg-Wastljung AC (2011) Polymorphism and divergence in two willow species, *Salix viminalis* L. and *Salix schwerinii* E. Wolf. G3 (Bethesda) 1:387–400.

Brereton NJB, Pitre FE, Hanley SJ, Ray MJ, Karp A, Murphy RJ (2010) QTL mapping of enzymatic saccharification in short rotation coppice willow and its independence from biomass yield. BioEnergy Res 3:251–261.

Brereton NJB, Pitre FE, Ray MJ, Karp A, Murphy RJ (2011) Investigation of tension wood formation and 2, 6-dichlorbenzonitrile application in SRC willow composition and enzymatic saccharification. Biotechnol Biofuels 4:13.

Brereton NJB, Ray MJ, Shield IF, Martin P, Karp, A, Murphy RJ (2012) Reaction wood – a key cause of variation in cell wall recalcitrance in willow. Biotechnol Biofuels 5:83.

Cerasuolo M, Richter GM, Cunniff J, Purdy S, Shield I, Karp A (2013) A pseudo-3D model to optimise the target traits of light interception in short-rotation coppice willow. Agric For Meteorol 173:127–138.

Cohen D, Bogeat-Triboulot M-B, Tisserant E et al. (2010) Comparative transcriptomics of drought responses in *Populus*: a meta-analysis of genome-wide expression profiling in mature leaves and root apices across two genotypes. BMC Genomics 11:630.

Felton J, Sundberg B (2013) Biology, chemistry and structure of tension wood. In: Fromm J (ed) Cellular aspets of wood formation. Plant Cell Monographs 20. Springer, Berlin.

Grattapaglia D, Resende MDV (2011) Genomic selection in forest tree breeding. Tree Genet Genomes 7:241–255.

Gullberg U (1993) Towards making willows pilot species for coppicing production. Forestry Chron 69:721–726.

Hanley SJ (2003) Genetic mapping of important agronomic traits in biomass willow. Ph.D. thesis. University of Bristol, UK.

Hanley S, Barker JHA, Van Ooijen JW, Aldam C, Harris SL, Ahman I, Larsson S, Karp A (2002) A genetic linkage map of willow (*Salix viminalis*) based on AFLP and microsatellite markers. Theor Appl Genet 105:1087–1096.

Hanley SJ, Mallott MD, Karp A (2006) Alignment of a *Salix* linkage map to the *Populus* genomic sequence reveals macrosynteny between willow and poplar genomes. Tree Genet Genomes 3:35–48.

Hanley SJ, Pei MH, Powers SJ, Ruiz C, Mallott MD, Barker JHA, Karp A (2011) Genetic mapping of rust resistance loci in biomass willow. Tree Genet Genomes 7:597–608.

Harfouche A, Meilan R, Kirst M, Morgante M, Boerjan W, Sabatti M, Mugnozza GS (2012) Accelerating the domestication of forest trees in a changing world. Trends Plant Sci 17:64–72.

Joosen RVL, Ligterink W, Hilhorst HWM, Keurentjes JJB (2009) Advances in genetical genomics of plants. Curr Genomics 10:540–549.

Karlberg A, Englund M, Petterle A, Molnar G, Sjodin A, Bako L, Bhalerao RP (2010) Analysis of global changes in gene expression during activity-dormancy cycle in hybrid aspen apex. Plant Biotechnol 27:1–16.

Karp A (2013) Willows as a source of renewable fuels and diverse products. In: Fenning T (ed.) Challenges and opportunities for the World's forests in the 21st century. Springer, Dordecht, The Netherlands (in press).

Karp A, Shield I (2008) Bioenergy from plants and the sustainable yield challenge. New Phytol 179:15–32.

Karp A, Edwards KJ, Bruford M et al. (1997) Molecular technologies for biodiversity evaluation: opportunities and challenges. Nat Biotechnol 15:625–628.

Karp A, Hanley SJ, Trybush SO, Macalpine W, Pei M, Shield I (2011) Genetic improvement of willow for bioenergy and biofuels. J Integr Plant Biol 53:151–165.

Karp A, Richter GM, Shield IF, Hanley SH (2013) Genetics, genomics and crop modelling: integrative approaches to the improvement of biomass willows. In: Carpita N, Buckeridge MS, McCann MC (eds) Plants and bioenergy. Chapter 7. Springer, New York (in press).

Köhler J, Baumbach J, Taubert J, Specht M, Skusa A, Rüegg A, Rawlings C, Verrier P, Philippi S (2006) Graph-based analysis and visualization of experimental results with ONDEX. Bioinformatics 22: 1383–1390.

Larsson S (1997) Commercial breeding of willow for short rotation coppice. Aspects Appl Biol 49:215–218.

Larsson S (1998) Genetic improvement of willow for short-rotation coppice. Biomass Bioenergy 15:23–26.

Larsson S (2001) Commercial varieties from the Swedish willow breeding programme. Aspects Appl Biol 65:193–198.

Ledin S (1998) Environmental consequences when growing short rotation forests in Sweden. Biomass Bioenergy 15:49–55.

Li D, Wang L, Wang M, Xu Y-Y, Luo W, Liu Y-J, Xu Z-H, Li J, Chong K (2009) Engineering OsBAK1 gene as a molecular tool to improve rice architecture for high yield. Plant Biotechnol J 7:791–806.

Lin J, Gunter LE, Harding SA, Kopp RF, McCord RP, Tsai CJ, Tuskan GA, Smart LB (2007) Development of AFLP and RAPD markers linked to a locus associated with twisted growth in corkscrew willow (*Salix matsudana* 'Tortuosa'). Tree Physiol 27:1575–1583.

Lindegaard KN, Barker JHA (1997) Breeding willows for biomass. Aspects Appl Biol 49:155–162.

Lindegaard KN, Parfitt RI, Donaldson G et al. (2001) Comparative trials of elite Swedish and UK biomass willow varieties. Aspects Appl Biol 65:183–192.

Liu E, Wang Y, Xu La, Huang M (2011) A genetic frame map of *Salix suchowensis* × *S. erioclada* based on SSR and SRAP markers. Sci Silvae Sin 47:23–30.

Love CG, Andongabo AE, Wang J, Carion PW, Rawlings CJ, King GJ (2012) InterStoreDB: a generic integration resource for genetic and genomic data. J Integr Plant Biol 54:345–355.

Luo CW, Tsementzi D, Kyrpides N, Read T, Konstantinidis KT (2012) Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. PLoS One 7: Article No. e30087.

MacAlpine WJ, Shield IF, Trybush SO, Hayes CM, Karp A (2008) Overcoming barriers to crossing in willow (*Salix* spp.) breeding. Aspects Appl Biol 90:173–180.

Margulies M, Egholm M, Altman WE et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376–380.

Marques WL, Salazar MM, Oliveira Camargo EL, Lepikson-Neto J, Tiburcio RA, do Nascimento LC, Guimaraes Pereira GA (2013) Identification of four *Eucalyptus* genes potentially involved in cell wall biosynthesis and evolutionarily related to SHINE transcription factors. Plant Growth Regul 69:203–208.

Munoz C, Baeza J, Freer J, Teixeira Mendonca R (2011) Bioethanol production from tension and opposite wood of *Eucalyptus globulus* using organosolv pretreatment and simultaneous saccharification and fermentation. J Ind Microbiol Biotechnol 38:1861–1866.

Newsholme C (1992) Willows. The genus *Salix*. Timber Press, Inc., Oregon, USA.

Olarte A, Mantri N, Nugent G, Wohlmuth H, Li CG, Xue C, Pang E (2013) A gDNA microarray for genotyping *Salvia* species. Mol Biotechnol 54:770–783.

Pei MH, Royle DJ, Hunter T (1996) Pathogenic specialization in *Melampsora epitea* var *epitea* on *Salix*. Plant Pathol 45:679–690.

Ramstedt M (1999*a*) *Melampsora* leaf rust: the most serious disease of *Salix* in SRF. Vaxtskyddsnotiser 63:27–32.

Ramstedt M (1999*b*) Rust disease on willows – virulence variation and resistance breeding strategies. For Ecol Manage 121:101–111.

Ramstedt M, Hurtado S, Astrom B (2002) Pathotypes of *Melampsora* rust on *Salix* in short-rotation forestry plantations. Plant Pathol 51:185–190.

Ray MJ, Brereton NJB, Shield I, Karp A, Murphy RJ (2012) Variation in cell wall composition and accessibility in relation to biofuel potential of short rotation coppice willows. BioEnergy Res 5:685–698.

Rönnberg-Wastljung AC, Tsarouhas V, Semerikov V, Lagercrantz U (2003) A genetic linkage map of a tetraploid *Salix viminalis* × *S. dasyclados* hybrid based on AFLP markers. For Genetics 10: 185–194.

Rönnberg-Wastljung AC, Glynn C, Weih M (2005) QTL analyses of drought tolerance and growth for a *Salix dasyclados* × *Salix viminalis* hybrid in contrasting water regimes. Theor Appl Genet 110:537–549.

Ronnberg-Wastljung AC, Ahman I, Glynn C, Widenfalk O (2006) Quantitative trait loci for resistance to herbivores in willow: field experiments with varying soils and climates. Entomol Exp Appl 118:163–174.

Ronnberg-Wastljung A-C, Samils B, Tsarouhas V, Gullberg U (2008) Resistance to *Melampsora larici-epitea* leaf rust in *Salix*: analyses of quantitative trait loci. J Appl Genet 49:321–331.

Samils B, Ronnberg-Wastljung A-C, Stenlid J (2011) QTL mapping of resistance to leaf rust in *Salix*. Tree Genet Genomes 7:1219–1235.

Sennerby-Forsse L, Zsuffa L (1995) Bud structure and resprouting in coppiced stools of *Salix viminalis* L, *S. eriocephala* Michx, and *S. amygdaloides* Anders. Trees Struct Funct 9:224–234.

Serapiglia MJ, Humiston MC, Xu H, Hogsett DA, de Orduna RM, Stipanovic AJ, Smart LB (2013) Enzymatic saccharification of shrub willow genotypes with differing biomass composition for biofuel production. Front Plant Sci 4:57–57.

Snoek LB, Terpstra IR, Dekter R, Van den Ackerveken G, Peeters AJM (2012) Genetical genomics reveals large scale genotype-by-environment interactions in *Arabidopsis thaliana*. Front Genet 3:317–317.

Stolarski MJ, Szczukowski S, Tworkowski J, Klasa A (2013) Yield, energy parameters and chemical composition of short-rotation willow biomass. Ind Crop Prod 46:60–65.

Stott KG (1984) Improving the biomass potential of willow by selection and breeding. In Perttu K (ed.) Ecology and management of forest biomass production systems. Swedish University of Agricultural Sciences, pp. 233–260.

Tsarouhas V, Gullberg U, Lagercrantz U (2001) An AFLP and RFLP linkage map and quantitative trait locus (QTL) analysis of growth traits in *Salix*. Theor Appl Genet 105:277–288.

Tsarouhas V, Gullberg U, Lagercrantz U (2003) Mapping of quantitative trait loci controlling timing of bud flush in *Salix*. Hereditas 138:172–178.

Tsarouhas V, Gullberg U, Lagercrantz U (2004) Mapping of quantitative trait loci (QTLs) affecting autumn freezing resistance and phenology in *Salix*. Theor Appl Genet 108:1335–1342.

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

Vanholme B, Cesarino I, Goeminne G et al. (2013) Breeding with rare defective alleles (BRDA): a natural *Populus nigra* HCT mutant with modified lignin as a case study. New Phytol 198:765–776.

Verwijst T (1996*a*) Cyclic and progressive changes in short-rotation willow coppice systems. Biomass Bioenergy 11:161–165.

Verwijst T (1996*b*) Stool mortality and development of a competitive hierarchy in a *Salix viminalis* coppice system. Biomass Bioenergy 10:245–250.

Volk TA, Abrahamson LP, Nowak CA, Smart LB, Tharakan PJ, White EH (2006) The development of short-rotation willow in the northeastern United States for bioenergy and bioproducts, agroforestry and phytoremediation. Biomass Bioenergy 30:715–727.

Vos J, Evers JB, Buck-Sorlin GH, Andrieu B, Chelle M, de Visser PHB (2010) Functional-structural plant modelling: a new versatile tool in crop science. J Exp Bot 61:2101–2115.

Ward S, Salmon J, Hanley S, Karp A, Leyser O (2013) Using *Arabidopsis* to study shoot branching in biomass willow (*Salix* spp.). Plant Physiol 162:800–811.

Wenger Y, Galliot B (2013) RNAseq versus genome-predicted transcriptomes: a large population of novel transcripts identified in an Illumina-454 Hydra transcriptome. BMC Genomics 14:204.

Yang XH, Kalluri UC, DiFazio SP, Wullschleger SD, Tschaplinski TJ, Cheng ZMM, Tuskan GA (2009) Poplar genomics: state of the science. Crit Rev Plant Sci 28:285–308.