Supplementary Methods

Construction of a high-density linkage map

To obtain informative markers that could be used in the creation of a linkage map, markers were filtered in several steps. First, the vcf file was filtered with `bcftools` (Narasimhan et al. 2016) to only include bi-allelic SNPs, without low quality tags and with a minor allele frequency (MAF) > 0.25 in the parents. All SNPs outside the extended probe regions (120 bp ± 100 bp) and SNPs having a genotype depth (DP) falling outside the range of 10-100 were removed. Progeny genotypes were then filtered using a custom awk script, retaining only genotype calls matching the possible variants available in the Punnet square based on parental genotypes. Genotypes that did not match this criterion were recoded as missing data. Genotyping information was then extracted from the vcf file and all remaining filtering steps were performed in `R` (R Core Team 2018).

For the map construction we only used markers where both genotyping methods in the parents (capture probes and WGS) showed concordance, where at least one of the parents was heterozygous and where no more than 20% of the progeny had missing data. A chi-square test for segregation distortion was performed on all remaining markers and all markers with a significance level > 0.005 were kept. Finally, only the best marker, in terms of lowest level of missing data and most balanced segregation pattern was kept for each probe. Genotypes and marker segregation pattern were then recoded to `BatchMap` input format (Schiffthaler et al. 2017). The resulting file contained 764 F_1 progeny and 19,520 probe markers, segregating either in the mother (cross type D1.10), the father (cross type D2.15) or both parents (cross type B3.7).

Framework linkage maps were created using `BatchMap` (Schiffthaler et al. 2017), a parallelized implementation of `OneMap` (Margarido et al. 2007), using the pseudo test cross strategy. Pairwise estimates of recombination frequency were calculated between all markers using a LOD score of 8 and a maximum recombination fraction (`max.rf`) of 0.35. To reduce

the number of redundant markers in the map, identical markers (showing no recombination events between them) were collected into bins where one representative marker (the marker with lowest amount of missing data) per bin was used in subsequent analyses. Following binning, markers were grouped into linkage groups (LGs) using a LOD threshold of 12 and further split into a maternal (D1.10 and B3.7 markers) and paternal (D2.15 and B3.7 markers) mapping population with B3.7 markers acting as a bridge between the two parental maps. Marker ordering along the LGs was estimated using 16 rounds of the `RECORD` (Recombination count and ordering) algorithm (Van Os et al. 2005) parallelized over 16 cores. Genetic distance estimates were calculated using three rounds of the 'map batches' approach (Schiffthaler et al. 2017) using the Kosambi mapping function. In successive rounds, markers were rippled in sliding windows of eleven, nine and seven markers, respectively, using 32 ripple cores and 2 phasing cores (Margarido et al. 2007, Schiffthaler et al. 2017).

A consensus map of the two parental framework maps was created with the R-package `LPmerge` (Endelmann et al. 2014) using a maximal interval setting ranging from one to ten and equal weight to the two parental maps. The consensus map with the lowest mean root mean square error (RMSE) was set as the best consensus map for each LG.

In order to estimate the correspondence between different LGs from the linkage map and chromosomes in *P. trichocarpa*, probe sequences were mapped against the masked *P. trichocarpa* genome assembly v3.0 using `BLASTn` (Altschul et al. 1990). For each probe with a corresponding marker in the consensus map, the BLAST hit with the highest bitscore value was considered to be the homologous region of the *P. trichocarpa* genome. The number of homologous regions observed between the *P. tremula* LGs and the *P. trichocarpa* chromosomes were used to assign *P. tremula* LGs to corresponding *P. trichocarpa* chromosomes (Figure S1).

Establishing relationships between genetic and physical positions

We used the Python software `AllMaps` (Tang et al. 2015) to create physical chromosomes from the *P. tremula* genome assembly v.1.1 based on the framework genetic maps. Briefly, `AllMaps` uses information from linkage maps to physically anchor scaffolds from a genome assembly into chromosomes. All markers that had been placed into bins at the beginning of the linkage map creation were reintroduced to the final parental framework maps by placing them at the same chromosome and genetic distance as the bin representative marker. All scaffolds in the framework maps that had markers mapped to more than one chromosome (340 scaffolds), or where markers were mapped to different positions on a single LG but more than 20 centiMorgans (cM) apart (19 scaffolds), were split and placed using the corresponding positions of the markers (Figures S2, S3 and S4). To achieve as accurate scaffold splits as possible, assembly gaps and gene annotations were considered. For each scaffold region anchored in the framework maps, the largest assembly gap outside gene models were chosen to split scaffolds. If no assembly gaps were present within the split region, the scaffolds were split in the middle of an intergenic region by artificially creating a gap of size 1 bp. However, if the split region was positioned within a single gene model, the gene models were split either at the largest assembly gap or by artificially creating a gap of size 1 bp in the middle of the region (Figure S5). The python software `jcvi` (Tang et al. 2015) was used to physically split and rename multi-mapped scaffolds. One scaffold, *Potra001073*, contained two markers that could not be split using the rules above since they were positioned in two overlapping gene models appearing on the same strand. These markers were therefore removed from the map. After splitting scaffolds, the linkage map marker positions were translated to the new split scaffold assembly positions using UCSC `liftOver` (Hinrichs et al. 2006) and used as input to `AllMaps`.

`AllMaps` was run according to instructions (https://github.com/tanghaibao/jcvi/wiki/ALLMAPS) using the parental framework maps, here

after referred to as the 'female' and 'male' map, respectively. The maps were merged into the input bed file and weighed equally (1) for scaffold ordering. After ordering, the built-in gap length estimation in `AllMaps` was run to produce more precise lengths for the chromosomes. The chromosome-scale assembly produced will be referred to as *P. tremula* v1.2.

Linkage map-based recombination rates

The parental framework maps as well as the consensus map were edited with custom awk scripts to match the input format specified in the manual of the R-package `MareyMap` (Rezvoy et al. 2007). All genetic maps were converted to bed format using a custom awk script for easy lift over to the new physical assembly with the UCSC `liftOver` tool (Hinrichs et al. 2006). The lift-over was performed with the '`–bedPlus`' option enabled to carry over extra columns and then recoded back to `MareyMap` input format. Some of the male chromosomes were reversed relative to the female and consensus maps (see negative ρ-values in Figure S6). This was done by taking the absolute values of the genetic distance column after subtracting the maximum value of genetic distance from all the values in the column using a custom-made Python script. The edited maps were read into `MareyMap` and two obvious outliers caused by a splitting oversight on chr5 (Figures S2 and S6) and an artificial gap caused by `LPmerge` when creating the consensus map from the parental framework maps in chr16 (Figures S2 and S6) were removed. Finally, we used the 'sliding window' method in `MareyMap` to estimate recombination rate in windows of 1Mbp, with a step size of 250 kbp and a minimum number of SNP's per window of 8, in order to avoid regions with large gaps being assigned recombination values.

LD-based recombination map

To estimate LD-based recombination rates, the vcf-file with data for the 94 SwAsp individuals from Wang et al. (2018) was lifted over to v1.2 genome coordinates by first recoding the file to bed format using `vcf2bed` from the `BEDOPS` toolkit (Neph et al. 2012), lifted over with UCSC

liftOver (Hinrichs et al. 2006) and finally recoded back to vcf format. The resulting vcf file was filtered using vcftools (Danecek et al. 2011), retaining only bi-allelic SNPs with a minor allele frequency greater than 0.05 (maf > 0.05) and that showed no evidence for deviations from Hardy-Weinberg equilibrium ($p > 0.002$).

We used LDhelmet v.1.10 (Chan et al. 2012) to produce a LD-based recombination map. LDhelmet handles a maximum of 25 diploid individuals (ie. 50 haplotypes), and we therefore sampled a random subset of 25 individuals from the 94 re-sequenced SwAsp individuals utilizing the vcftools '--max-indv' -option. The subset vcf was split into separate files according to chromosomes to avoid memory issues when running LDhelmet. Full FASTA-files were produced for the 25 individuals by reassigning SNP-positions at the corresponding sites within the reference FASTA for *P. tremula* v1.2 using the 'vcf-consensus' script from vcftools. This was done twice per individual to produce both copies of the chromosome. Individual files were then concatenated together to generate a single FASTA file per chromosome with data for all 25 individuals.

The LDhelmet preparatory files were produced as suggested in the LDhelmet v1.10 manual (Chan et al. 2012). We produced Padé coefficients and lookup tables separately for each of the 19 chromosomes. The configuration files were created using a window size of 50 and the lookup table was produced using the recommended values found in the LDhelmet manual (-t 0.01 -r 0.0 0.1 10.0 1.0 100.0). Similarly, the Padé coefficients tables were produced using the recommended values from the manual (-t 0.01 -x 11). To further lighten the computational load during the recombination estimation step, we opted to use the SNPs and pos files for our chromosomes. These files were produced for each chromosome separately using the '--ldhelmet' -option for vcftools v0.1.15. We used the recommended values from the LDhelmet manual as *Populus tremula* has similar levels of nucleotide polymorphism and

extent of linkage disequilibrium as *Drosophila melanogaster*, on which the recommended settings in `LDhelmet` are based upon (Wang et al. 2016)

Methylation sequencing

Methylation levels were estimated using bisulfite sequencing data from six SwAsp individuals. The six individuals were bisulfite sequenced using two biological replicates per individual using paired-end (2x150) sequencing on an Illumina HiSeq X at the National Genomics Infrastructure facility at Science for Life Laboratory in Uppsala, Sweden. Based on recommendation from the NGI facility, samples were sequenced to an average depth of 60x as approximately 50% of the bisulfite sequencing data cannot be mapped uniquely due to excessive damage induced by the bisulfate treatment. The raw bisulfite-sequencing reads were trimmed using `trimGalore` v. 0.4.4 (https://github.com/FelixKrueger/TrimGalore), a wrapper around `Cutadapt` (Martin 2011) and `FastQC` (Andrews 2010), with a paired-end trimming mode and otherwise default settings. In order to obtain as accurate methylation calls as possible, polymorphic substituted versions of the *P. tremula* v1.1 assembly (Lin et al. 2018) were created for each sample separately using `bcftools` consensus (Narasimhan et al. 2016) and bisulfite converted and indexed with `bismark_genome_preparation` (Krueger and Andrews 2011). Trimmed reads for all individuals were mapped against the corresponding converted reference genomes using `Bismark` (Krueger and Andrews 2011) and `Bowtie2` (Langmead and Salzberg 2012). In order to remove optical duplicates from the BAM files, we ran `deduplicate_bismark` with default settings before methylation levels were extracted using `bismark_methylation_extractor` with the following settings: `--bedGraph --gzip --comprehensive --CX --scaffolds --buffer size 20G`, which produced both context files and coverage files. For one of the sequenced individuals (SwAsp046), subsequent analyses suggested that the two biological replicates were actually derived from two genetically distinct individuals, likely due to a sampling mix-up in the

common garden. For all individuals, the biological replicate with largest amount of data available was used in downstream analyses. The `Bismark` results were lifted over to v1.2 coordinates using UCSC `liftOver` with the '`--bedPlus`' option. Coverage-files were filtered for low (< 5) and high (> 44) coverage observations to remove spurious results due to low coverage or collapsed duplicate genomic regions, respectively. Following filtering, coverage files for all six samples were merged and the different contexts of methylation (GpG, CHG and CHH) were extracted from the `Bismark` context files.

Literature cited

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403-410.

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Booker TR, Ness RW, Keightley PD. 2017. The recombination landscape in wild house mice inferred using population genomic data. *Genetics* **207**: 297–309.

Chan AH, Jenkins PA, Song YS. 2012. Genome-wide fine-scale recombination rate variation in *Drosophila melanogaster*. *PLoS Genetics* **8**: e1003090.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. *Bioinformatics* **27**: 2156–2158.

Endelman JB, Plomion C. 2014. LPmerge: An R package for merging genetic maps by linear programming. *Bioinformatics* **30**: 1623–24.

Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, Diekhans M, Furey TS, Harte RA, Hsu F, et al. 2006. The UCSC Genome Browser Database: update 2006. *Nucleic Acids Research* **34**: D590–8.

Krueger F, Andrews SR. 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**: 1571-1572.

Lin Y-C, Wang J, Delhomme N, Schiffthaler B, Sundström G, Zuccolo A, Nystedt B, Hvidsten TR, de la Torre A, Cossu RM, et al. 2018. Functional and evolutionary genomic inferences in *Populus* through genome and population sequencing of American and European aspen. *Proceedings of the National Academy of Sciences of the United States of America* 115: E10970–E10978.

Langmead B, Salzberg S. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**: 357-359.

Margarido GRA, Souza AP, Garcia AAF. 2007. OneMap: Software for Genetic Mapping in Outcrossing Species. *Hereditas* **144**: 78–79.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**: 10-12.

Narasimhan V, Danecek P, Scally A, Xue Y, Tyler-Smith C, Durbin R. 2016. BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics* **32**: 1749–1751.

Neph S, Kuehn MS, Reynolds AP, Haugen E, Thurman RE, Johnson AK, Rynes E, Maurano MT, Vierstra J, Thomas S, et al. 2012. BEDOPS: high-performance genomic feature operations. *Bioinformatics* **28**: 1919–1920.

R Core Team. 2018. R: A language and environment for statistical computing. *R Foundation for statistical Computing*, Vienna, Austria.

Rezvoy C, Charif D, Guéquen L, Marais GAB. 2007. MareyMap: an R-based tool with graphical inter-face for estimation recombination rates. *Bioinformatics* **23**: 2188-2189.

Schiffthaler B, Bernhardsson C, Ingvarsson PK, Street NR. 2017. BatchMap: A parallel implementation of the OneMap R package for fast computation of F1 linkage maps in outcrossing species. *PLoS ONE* **12**: e0189256.

Tang H, Zhang X, Miao C, Zhang J, Ming R, Schnable JC, Schnable PS, Lyons E, Lu J. 2015. ALLMAPS: robust scaffold ordering based on multiple maps. *Genome Biology* **16**: 3.

Van Os H, Stam P, Visser RGF, Van Eck HJ. 2005. RECORD: A novel method for ordering loci on a genetic linkage map. *Theoretical and Applied Genetics* **112**: 30–40.

Wang J, Street NR, Scofield DG, Ingvarsson PK. 2016. Natural selection and recombination rate variation shape nucleotide polymorphism across the genomes of three related *Populus* species. *Genetics* **202**: 1185–1200.