

Linkage Mapping of Biomass Production and Composition Traits in a *Miscanthus Sinensis* Population

Raphael RAVERDY

INRAE - Unité transfrontalière BioEcoAgro

Kristelle LOURGANT

INRAE - Unité transfrontalière BioEcoAgro

Emilie MIGNOT

INRAE UE GCIE Picardie

Stéphanie ARNOULT

INRAE UE GCIE Picardie

Guillaume BODINEAU

INRAE UE GBFor

Yves GRIVEAU

IJPB: Institut Jean-Pierre Bourgin

Cristiane H. TANIGUTI

University of Sao Paulo: Universidade de Sao Paulo

Maryse BRANCOURT-HULMEL (✉ Maryse.Hulmel@inrae.fr)

BioEcoAgro Cross-border: Transfrontalière BioEcoAgro <https://orcid.org/0000-0002-8679-7537>

Research Article

Keywords: integrated genetic map, QTL mapping, orthologous genes, age effect, climatic condition effect

Posted Date: September 3rd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-751573/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at BioEnergy Research on March 10th, 2022.
See the published version at <https://doi.org/10.1007/s12155-022-10402-8>.

Abstract

Breeding miscanthus for biomass production and composition is essential for targeting high-yielding genotypes suited to different end-uses. Our objective was to understand the genetic determinism of these traits in *M. sinensis*, according to different plant ages and environmental conditions. A diploid population was established in two locations according to a staggered-start design, which made the “year” effect partitioned into “age” and “growing season” effects. An integrated genetic map of 2,602 SNP markers distributed across 19 LGs, was aligned with the *M. sinensis* reference genome and spanned 2,770 cM. The QTL mapping was based on Best Linear Unbiased Predictions estimated across three climatic conditions and at least three ages in both locations. 260 and 283 QTL were related to biomass production and composition traits, respectively. In each location, 40%-60% were related to biomass production traits and stable across different climatic conditions and ages, and 30% to biomass composition traits. Ten to fifteen% were stable for both trait types across locations. Twelve QTL clusters were established based on either biomass production or composition traits, and validated by high genetic correlations between the traits. Sixty-two putative *M. sinensis* genes, related to the cell-wall, were evidenced in the QTL clusters of biomass composition traits, and orthologous to those of sorghum and maize. Twelve of them were differentially expressed and belonged to gene families related to the cell-wall biosynthesis identified in other miscanthus studies. These stable QTL constitute new insights into Marker-Assisted Selection breeding while offering a joint improvement of biomass production or composition traits.

Introduction

Miscanthus is a perennial C4 crop that produces valuable lignocellulosic biomass, mainly for bioenergy end-uses, bio-based products and animal bedding [1–5]. However, only one clone from *Miscanthus × giganteus* interspecific hybrid ($2n = 3x = 57$) is currently available for commercialization in Europe and the United States [6, 7]. It originates from a natural interspecific cross between a tetraploid *M. sacchariflorus* ($2n = 4x = 76$) and a diploid *M. sinensis* ($2n = 2x = 38$) [8] and obtains high biomass yields under various environmental conditions [9, 10]. As *M. × giganteus* is sterile, the potential risk of invasiveness from the seeds spreading to a new environment is avoided, but this hampers the breeding of the crop [11, 12]. Moreover, such narrow genetic variability can be risky in case of pest adaptation, and the corresponding phenotypic variability may not be sufficient for the different end-use requirements [8]. The two parents of *M. × giganteus*, originating from East Asia, present high genetic variability and are adapted to extended environmental conditions [13–16]. For example, *M. sinensis* reaches the same amount of biomass production as *M. × giganteus* under specific conditions [17]: this makes it an interesting candidate for breeding new varieties at an intraspecific level, or for creating new *M. × giganteus* clones at an interspecific level, despite its self-incompatibility [18]. Breeding miscanthus aims at improving biomass production and composition traits by creating new cultivars, adapted to a range of different environments. However, the optimal time to collect reliable phenotypic data for these traits is a major point of concern in miscanthus breeding. Indeed, the yield-building phase of the plants can last three

years after crop establishment, and even this can vary between miscanthus species [19]. The plants then reach a yield-plateau phase during which biomass production is more stable [20]. Selecting plants only on their phenotype during the yield-building phase may thus be unreliable (Raverdy et al., submitted to BioEnergy Research), especially since the different progenies may reach their full growth potential at different times. Marker-assisted selection (MAS) programs would thus be a considerable tool for improving the breeding efficiency in miscanthus, as they would make it possible to select young plants on the basis of their genetic information, without waiting for them to reach the yield-plateau phase to record their biomass production phenotype. Such programs require the construction of genetic maps in order to carry out QTL mapping.

Because of its self-incompatibility, miscanthus is an outcrossing species with a high level of heterozygosity and therefore, the genetic mapping methods that were initially developed based on inbred line populations have to be adapted to full-sib (F1) populations, as is also the case for other perennial crops such as sugarcane or rubber tree [21–23]. Full-sib populations have a maximum of four different segregating alleles per locus, which makes the haplotype phase estimation more complex than for inbred lines populations, for which a maximum of two alleles segregate.

The initial traditional method for building genetic maps in full-sib populations was the “pseudo-testcross” strategy [24]. It only considered markers that segregate in a 1:1 ratio, which carry recombination information separately for each parent, thus generating separate linkage maps. Later, Maliepaard et al. [25] revealed maximum likelihood estimators for all possible segregation patterns in the population, such as the 3:1, 1:2:1 and 1:1:1:1 ratios. Wu et al. [26, 27] included the estimators in a multipoint approach that is capable of estimating the phase and the genetic distance simultaneously by using information from the entire linkage group and generating integrated genetic maps. The approach described in Wu et al. [26, 27] was implemented in the OneMap package [28, 29].

Compared with “pseudo-testcross” linkage maps, integrated genetic maps provide a more accurate estimation of linkage distances and phases between markers, as well as of locations of the Quantitative Trait Loci (QTL), once they produce more saturated linkage maps and focus on estimated progeny haplotypes. Moreover, Gazaffi et al. [30] put forward a model for carrying out QTL mapping with an integrated genetic map, using a Composite Interval Mapping (CIM) model for outcrossing species. In this model, three orthogonal contrasts are estimated based on maximum likelihood and mixture models, which are then used to estimate additive and dominance effects, QTL segregation patterns and linkage phases within the markers.

The initial miscanthus genetic map was developed by Atienza et al. [31] based on a *M. sinensis* population, by using Random Amplified Polymorphic DNA (RAPD) markers, and was composed of 28 linkage groups (LGs). This unsaturated map was used to map QTL for morphological traits, biomass yield and combustion quality [32–35]: however, the 28 LGs that constitute the genetic map, compared to the base chromosome number in miscanthus ($x = 19$), as well as the small population size of 89 F1 individuals, make it difficult to compare the QTL with recent studies in which saturated maps are

presented with the expected number of 19 LGs. Ma et al. [36] created a high-resolution linkage map of *M. sinensis* and identified 19 LGs for the first time, which corresponded to the base chromosome number of the species. Gifford et al. [37] identified 72 QTL associated with biomass productivity, by using the integrated genetic map of *M. sinensis* constructed by Swaminathan et al. [38], which was also made up of 19 LGs. Dong et al. [39] developed six high-density parental genetic maps using the pseudo-testcross strategy and a consensus map which integrated *M. sinensis* and *M. sacchariflorus*. Using these maps derived from three interconnected miscanthus populations, they compared four QTL analysis methods to detect marker-trait associations for 14 biomass production traits: they identified 109 to 288 QTL which mapped into 86 to 157 meta-QTL, depending on the method used. Van der Weijde et al. [40] constructed two parental maps of *M. sinensis* according to the pseudo-testcross strategy and identified 86 QTL related to biomass composition and conversion efficiency traits. Most of these studies identified QTL of *M. sinensis* biparental populations derived from crosses between different heterozygous parents, and these populations were evaluated during two years in one location. In apple trees, another perennial species, Segura et al. [41] carried out QTL mapping on an F1 progeny and demonstrated that the QTL detected are related to genetic, ontogenetic and climatic factors: this was possible by a staggered-start design [42–44], which made it possible to partition the year effect into age and growing season effects. The growing season effect itself corresponded to soil and climatic condition effects. These statements thus suggest that the number, position and effects of QTL for a trait of interest are determined according to the genetic variability available in the population studied, the methodology used to construct the corresponding genetic map, the experimental design choice and the conditions under which the plants grew. These conditions are inherent to the age of the plant, and the climatic and soil conditions for successive years.

In a previous study, we estimated the heritability values and genetic and phenotypic correlations of a *M. sinensis* population, for which we highlighted moderate to high heritability values for biomass production and composition traits, considering both age and climatic condition effects, and based on a staggered-start design carried out in two contrasted locations (Raverdy et al., submitted in BioEnergy Research). We thus expected the phenotypic data of this population to be relevant for undertaking QTL mapping. The objectives of this study were then: (1) to construct an integrated linkage map based on Single Nucleotide Polymorphism (SNP) markers according to the miscanthus reference genome, and (2) to detect QTL for biomass production and composition traits and identify stable QTL, while considering the range of years (*i.e.* climatic conditions), ages and locations evaluated. In the present study, the same population has therefore been analyzed, based on the staggered-start design that was established in two contrasted locations. As the stands were staggered twice in a single field in both locations, the growing season effects corresponded to climatic condition effects. To reach these goals, a Genotyping-By-Sequencing (GBS) approach was initially used to discover SNP markers according to the alignment with the miscanthus reference genome that was released in 2017 [45], and to construct an integrated linkage map for the population. QTL were then detected for both biomass production and composition traits over five consecutive years. To our knowledge, it is the first miscanthus genetic map that considers the alignment with the miscanthus reference genome and the first miscanthus study for which the QTL of biomass

production and composition traits are jointly detected for more than two years and in two contrasted locations. This will make it possible to assess marker-trait associations during and after the theoretical establishment phase of miscanthus, and according to different ages and climatic conditions. Based on the staggered-start design per location, the “year” effect was partitioned into “age” and “climatic condition” effects, which made it possible to detect QTL related to such conditions for the first time in miscanthus.

Materials And Methods

Mapping population and experimental design

Two diploid ornamental *M. sinensis* cultivars, “Malepartus” (Mal) and “Silberspinne” (Sil), were crossed in order to get an F1 full-sib progeny for mapping studies. These two ornamental parents putatively originate from central to southern Japan [15, 16]. The population of 248 F1 full-sib genotypes was obtained at the end of the germination phase and was thus available for genotyping. However, only 157 genotypes were finally established in the field, because some of the seedlings could not adapt to these growing conditions.

The F1 full-sib progeny and the two parental genotypes were cultivated with single plants, in two locations in France. These locations presented contrasted soil and climatic conditions. Experimental design in both locations was based on a staggered-start design [42]. One of the experiments was established at the “GCIE” INRAE (National Institute for Research on Agriculture, food, and Environment) experimental unit of Estrées-Mons (49°53'N, 3°00'E) in a deep loam soil (Orthic Luvisol according to the World Reference Base for Soil Resources, WRB). The other experiment was established at the “GBFOR” INRAE experimental unit of Orléans (47°49'N, 1°54'E) in a sandy soil (Dystric Cambisol, WRB). Each staggered-start design was made up of two stands or groups of genotypes that were organized in two plots established in two successive years in the same field: the first group of genotypes (G1) was established in 2014 while the second group (G2) was established in 2015 (Fig. 1). In each location, each plot was adjacent to the other plot in the field and separated by a border row: soil conditions were therefore similar between the two groups. In Estrées-Mons, 157 genotypes and the two parents were cultivated, with 82 genotypes common to each group (G1 and G2). In Orléans, 104 genotypes and the two parents were cultivated, with 59 genotypes common to G1 and G2 (Fig. 1). The number of genotypes between both groups (G1 and G2) was unbalanced due to the recalcitrance of some genotypes concerning the propagation and establishment steps. In both locations and each plot that corresponded to each group of genotypes sequentially established, single-plants were organized in an incomplete randomized block design [46] with five blocks. The genotypes were thus replicated in four of the five blocks on average, excepting the two parents which were replicated in all blocks of the two stands. Plant density was 1 plant per m², with single-plants equally spaced 1 meter apart within and between rows. For both locations, more details about the experimental design and the climatic conditions that correspond to the plant cycle are available in Raverdy et al. (submitted in BioEnergy Research).

Development of single nucleotide polymorphism markers

The genomic DNA of the 248 individuals of the progeny and of the two parents was extracted from seedlings at the INRAE Gentyane Genomic Service platform (Clermont-Ferrand, Puy-de-Dôme, France), using a sbeadex™ livestock kit (LGC Group, United Kingdom). A Genotyping-By-Sequencing (GBS) approach was used to discover Single Nucleotide Polymorphism (SNP) markers which corresponded to the population, and was carried out according to the protocols of Elshire et al. [47] and Cormier et al. [48], at the CIRAD (French Agricultural Research Centre for International Development) Genotyping platform core facility (GPTR, Montpellier, Hérault, France). The 96-plex libraries were prepared by digestion of the 250 DNA extractions using the *Pst*I restriction enzyme. The single-end sequencing was then carried out in three lanes on a HiSeq™ 3000 sequencer (Illumina® Inc., San Diego, CA, USA), at the “GenoToul GeT” platform (Auzerville, Haute-Garonne, France). The quality check of the reads was conducted using the “FastQC” software v.0.11.2 [49].

The “TASSEL 5 GBS” pipeline [50] was used in order to analyze the sequencing data. The raw reads of the individuals were initially grouped in tags. By using the “Bowtie 2” v.2.3.2 software [51], tags with a minimum count of 10 reads were retained and aligned with the *M. sinensis* reference genome sequence that was released in December, 2017 [45]. The resulting Variant Call Format file (VCF) which contains the markers and information regarding the individuals was then filtered using the “vcf2pop.1.0.py” software [52], based on a minor allele frequency (MAF) threshold of 0.05 and a maximum of 25% missing data per marker. Three marker types were available for the genetic map construction: markers that were heterozygous in both parents (ab × ab), called “Bridge” (Bri) markers and which segregated in a 1:2:1 ratio (aa, ab, bb), “Mal” markers that were heterozygous in the Malepartus parent and homozygous in the Silberspinne parent (ab × aa), and finally “Sil” markers that were homozygous in Malepartus and heterozygous in Silberspinne (aa × ab). Both of these last marker types segregated in a 1:1 ratio (aa, ab). According to the classification of Wu et al. [26], “Bri” markers were named B3.7, “Mal” markers were named D1.10 and “Sil” markers were named D2.15.

Genetic map construction

An integrated genetic map was constructed based on the mapping population, by using the “OneMap” R package [28, 53]. First, the redundant markers were removed from the analysis, and the remaining markers were tested according to Mendelian expected segregation by using a chi-square test and 0.05 probability for the alpha related to Bonferroni’s multiple test correction. The recombination fraction between all pairs of markers was then determined according to two-point tests [26]. The markers were then grouped based on their position on the reference genome chromosomes and a maximum recombination frequency of 0.35. Based on this grouping, 19 linkage groups (LGs) were obtained, which corresponded to the base chromosome number of miscanthus ($x = 19$). A high homology was highlighted between the markers grouped in each LG and their original position in accordance with the alignment with the reference genome chromosomes. For example, 90% of the markers grouped in LG1 were initially

identified in correspondence with chromosome 1 of the miscanthus reference genome sequence. The marker grouping for each of the 19 LGs was thus refined, by only keeping the markers that were in accordance with the corresponding chromosome of the miscanthus reference genome, which ensured that each marker was grouped in the right LG.

The segregation-distorted markers that were kept for the optimization of the marker grouping step were then discarded for marker ordering and phasing, to satisfy the QTL mapping model assumptions of Mendelian segregation. Marker ordering was then tested according to three different methods, and the Kosambi mapping function was used [54]. These different marker ordering methods were based either on heuristic algorithms or physical positions within the miscanthus reference genome. They had to be tested independently in order to find the best marker order among them. The best order was defined with the inspection of the expected pattern in the resulting recombination fraction matrix between markers, visualized in heatmaps. The ordering methods have already been investigated in different studies, with the aim of getting the best marker order, yielded either by the marker ordering the algorithms, or the physical positions within a reference genome [55, 56].

The first marker ordering method consisted in the ordering of the most informative markers (1:2:1) using an exhaustive search tool, which consisted in comparing all possible orders, and the remaining markers were positioned according to the “TRY” algorithm [57]. The resulting marker order was unsatisfactory according to the heatmaps (data not shown), although the “RIPPLE” algorithm [57] was used in order to improve it. A second marker ordering method was thus tested based on the Multi-Dimensional Scaling (MDS) method [58, 59] that was implemented in the OneMap software [28, 53]: although it improved the marker order, this approach did not provide the means to get a satisfactory marker order (data not shown). However, this approach made it possible to refine marker filtering, by removing some markers according to the Principal Curves method from the “MDSMap” R package [58, 59]. Finally, the third marker ordering method consisted in ordering the markers according to their positions identified within the miscanthus reference genome and to adjust the marker order based on recombination fractions between the markers. This final method was retained as it yielded the best marker order quality among the three methods.

Once the ordering was defined, the genetic distance was estimated based on multipoint approaches using Hidden Markov models [57], which consider outcrossing species [27]. In order to manage the presence of genotyping errors, as often encountered in mapping studies [55, 60–62], a genotyping error probability of 5% was considered in the Hidden Markov Model emission function implemented in the OneMap software, which made it possible to consider uncertainties between observed and estimated genotypes [63].

Phenotypic data analysis of biomass production and composition traits

Five biomass production traits and five biomass composition traits (expressed as a percentage of Dry Matter, %DM, or Cell-wall content, %CW) were studied. The phenotyping data were acquired over five

successive years in Estrées-Mons and four successive years in Orléans (Fig. 2). In order to name each trait in a relevant manner, a miscanthus ontology was developed at the INRAE BioEcoAgro research unit of Estrées-Mons by using the GnpIS multispecies integrative information system from the INRAE URGI of Versailles [64]. Four morphological traits were evaluated: canopy height (CH_cm), plant maximum height (HMax_cm), plant stem number (PSNb) and plant circumference (C50_cm). The aboveground biomass yield (ABM_tDMha) was measured after the winter harvest in late February and was expressed as tDM/ha. The biomass composition related traits were assessed based on Near-Infrared Spectroscopy (NIRS) predictions for all the plants of the population, using a set of calibration samples for which the composition traits were assessed using the Van Soest method [65]. These traits were Neutral Detergent Fiber, which corresponds to the cell-wall content (CW) as a percentage of dry matter (NDF_%DM), Acid Detergent Fiber (ADF_%DM), Cellulose (CL_%DM, CL_%CW), Hemicelluloses (HEM_%DM, HEM_%CW) and Acid Detergent Lignin (ADL_%DM, ADL_%CW). The phenotyping methods of all these traits are detailed in Raverdy et al. (submitted to BioEnergy Research).

For each location considered separately, the staggered-start design made it possible to analyze the phenotyping data by partitioning the “year” effect into “age” and “climatic condition” effects, according to two Linear Mixed Models [66]. An initial model (Model 1) was therefore assessed for each growing season or year (*i.e.* climatic condition), which took into account age effect modeling, and a second model (Model 2) was assessed for each age, which considered climatic condition effect modeling (Fig. 2). The two models were assessed using the Restricted Maximum Likelihood approach (REML), known to be suitable for analyzing unbalanced datasets, and the corresponding function was implemented in the “breedR” R package [67]. Model 1 was determined as:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$$

1

in which Y_{ijk} represents the phenotypic value measured on plant k of genotype i at age j ; μ is the overall mean; α_i is the random effect of genotype i ; β_j is the fixed effect of age j ; $(\alpha\beta)_{ij}$ is the random interaction between genotype i and age j and ϵ_{ijk} is the random residual for plant k of genotype i at age j . In addition, an autoregressive spatial component was included in the model using x and y coordinates in the field, in order to partition the residual ϵ_{ijk} into a spatially dependent parameter, θ_{ik} , for plant k of genotype i , and an independent remaining residual [68].

Model 2 was determined as:

$$Y_{ikl} = \mu + \alpha'_i + \gamma_l + (\alpha'\gamma)_{il} + \epsilon'_{ikl}$$

2

where each term is similar to those of Model 1, excepting age effect β_j which is replaced by climatic condition effect γ_l of year l , and the interaction between genotype i and age j $(\alpha\beta)_{ij}$ which is replaced by

the interaction between genotype i and climatic condition effect of year l (α'_{il}). The terms α'_i , ϵ'_{ikl} and θ'_{ik} were different from the effects given by the previous Model 1 because they were estimated based on different subsets of the data.

Block and spatial effects were both initially included in each model. As the models that only considered the spatial effect yielded the best Akaike Information Criterion (AIC) [69], the model was simplified and the block effect was finally not considered. It can be noted that the low performance of a statistical model that includes a block effect reinforced the similarity in the soil conditions between both stands of the staggered-start design. Based on each model described above, the Best Linear Unbiased Predictors (BLUP) of the random genotype (G) effect were estimated in order to carry out the Quantitative Trait Loci (QTL) mapping of miscanthus biomass production and composition traits. For each staggered-start design per location, both models were based on different subsets of the data as shown in the corresponding illustrations (Fig. 2) and the genotypes of both groups (G1 and G2) were considered (Fig. 1). For each location, the BLUP of the G effect that were estimated using Model 1 were related to the climatic condition of each studied year and were considered as independent from the age effect between G1 and G2 groups, *i.e.* represented by the fixed effect of the age and the random interaction between genotype and age (Fig. 2). Reciprocally, the BLUP of the G effect estimated using Model 2 were related to each age and were considered as independent from the climate effect between G1 and G2 groups, *i.e.* represented by the fixed effect of the climate and the random interaction between genotype and climatic conditions (Fig. 2). For biomass production and composition traits, the G effects estimated based on these two models were previously shown to account for the highest part of the variance components analyzed according to the staggered-start design in each location (Raverdy et al., submitted to BioEnergy Research).

For each trait and each condition, the distribution of the BLUP along with the parental values and phenotypic means were observed: as the data were normally distributed, no data transformation was needed (Figs. S5a, S5b, S5c, S5d and S5e).

Quantitative trait loci mapping

The Best Linear Unbiased Predictions (BLUP) of the five biomass production traits and the five biomass composition traits (expressed as %DM or %CW) previously cited were used for QTL mapping. Each condition of the staggered-start design per location was considered separately (Fig. 2). As miscanthus is an outcrossing species, a specific CIM model adapted to full-sib progeny was carried out by using the “fullsibQTL” R package [29, 30]. Outcrossing segregation patterns were considered in the model based on a multipoint approach which estimated three genetic effects according to three contrasts: two contrasts concerned the additive effects of the QTL alleles for each parent, and the third one was related to the intra-locus interaction (dominance) between additive effects on each parent. The conditional multipoint probabilities of QTL genotypes were obtained for every 1 cM interval. For each linkage group, the cofactors were selected using multiple linear regression with a stepwise procedure. The associated model selection was based on the AIC [69]. The QTL were defined for a threshold based on a 5% significance

level across distributed LOD scores that were obtained by selecting the second LOD profile peak from 1,000 permutations [70]. However, although some QTL were defined according to the same method, the corresponding threshold was based on a 10% significance level in order to detect a supplemental set of stable QTL between conditions (Fig. S1). The QTL concerned by this methodology were displayed with a “*” in the results section (Fig. 5). The three genetic effects previously detailed, the linkage phases, segregation patterns and the proportion of genotypic variation explained by each QTL (R^2) were estimated based on the CIM model. The QTL confidence intervals were calculated using the 2-LOD drop-off method [71].

As QTL mapping was carried out for each climatic condition and each age in each location separately, it was possible to highlight QTL that were stable according to these conditions. The QTL were defined as stable for a given trait when they co-localized under at least two conditions (Fig. 3). QTL could be stable across different climatic conditions in each location, as for example, a QTL detected in 2017 and 2018 in Estrées-Mons. QTL could also be stable across different ages in each location, if for example a QTL was detected at age 3 and age 4 in Estrées-Mons. Finally, QTL could also be stable across climatic conditions and ages in a given location: for example, when a QTL was detected in 2017 and at age 4 in Estrées-Mons. Finally, stable QTL were also identified when they co-localized under at least two conditions: (1) in both locations and under different climatic conditions: for example, a QTL found in Estrées-Mons in 2016 and in Orléans in 2018, (2) in both locations and for different ages : for example, a QTL found in Estrées-Mons at age 1 and in Orléans at age 2, (3) and lastly, in one location under a given climatic condition and in the other location for a given age: for example, a QTL found in Estrées-Mons in 2017 and in Orléans at age 3.

For each trait, the proportion of stable QTL across climatic conditions, ages and/or locations was then determined: in Estrées-Mons for example, the number of stable QTL that were found under at least two climatic conditions were divided by the total number of QTL found across the different years in Estrées-Mons, in order to establish the corresponding percentage.

QTL clusters were identified for a given climatic condition or a given age in each location. These clusters corresponded to the co-localization of at least three QTL for different traits: the biomass composition traits that were either expressed as %DM or %CW, were not considered as different traits. QTL clusters were made up of biomass production traits, biomass composition traits as well as both biomass production and composition traits. The reliability of these QTL clusters was verified according to the Pearson correlation coefficients computed based on BLUP, by using the “stats” R package, and visualized using the “corrplot” R package [72].

Identification of putative cell-wall-related genes in *M. sinensis*

Putative cell-wall-related genes were identified in *M. sinensis* based on the orthological relationships between *M. sinensis* and sorghum, and between *M. sinensis* and maize. These two species are indeed relatives of miscanthus and can be considered as genetic models for miscanthus, which makes it possible to take into account the genetic knowledge that is currently available for maize and sorghum.

Orthologous *M. sinensis* genes were initially identified based on two cell-wall-related gene lists, hereafter referred to as search lists, that were composed of 2,148 candidate genes for sorghum and 2,470 for maize (Virilouvet, personal communication). Secondly, these *M. sinensis* genes that are located between the flanking marker positions of each QTL cluster were selected based on the annotation file *Msinensis_497_v7.1.gene.gff3* from Phytozome, the Plant Comparative Genomics portal of the Department of Energy's Joint Genome Institute (<https://phytozome.jgi.doe.gov>) which contains 67,789 genes. Among these positional *M. sinensis* genes, the genes that were orthologous to cell-wall-related genes in sorghum and maize could finally be identified by comparison with the initial search lists.

Results

Single nucleotide polymorphism calling and filtering

The sequencing of Genotyping-By-Sequencing (GBS) libraries produced 1,161,537,843 raw reads based on a read length of 150 pb. Nineteen individuals out of the 248 were removed from the analysis, due to the low quality of the corresponding sequencing data: a total of 229 individuals was thus considered for the genetic map construction. The single nucleotide polymorphism (SNP) calling and the alignment with the *M. sinensis* reference genome made it possible to identify a final set of 149,741 biallelic SNP markers. The filtering of these markers then provided a selection of 9,330 high-quality SNP markers available for genetic mapping. The details of the markers according to their segregation type and coding [26] is available in Table 1.

Genetic map construction

The grouping of the SNP markers resulted in 3,774 SNP markers distributed across 19 LGs. Based on the initial marker dataset, redundant markers were removed, and markers unlinked to any of the 19 LGs were thus not considered for the analysis. It can be noted that LG12 was in fact partitioned into two LGs, because it was not possible to group all three marker types together due to the presence of only four B3.7 available markers (*i.e.* heterozygous in both parents): for that reason, LG12a was made up of “Bri” and “Mal” markers on the one hand, while LG12b was made up of “Bri” and “Sil” markers on the other. Finally, all LGs corresponded to the 19 chromosomes of the miscanthus reference genome, when checking each marker provenance according to their physical position in relation to the original alignment step. Segregation-distorted markers were initially kept in order to optimize the grouping phase, but they were then removed before the ordering step: 3,262 non-distorted SNP markers thus constituted the 19 LGs. Moreover, a supplemental analysis using the Principal Curves method led to the removal of an extra-set of 290 markers. After the ordering of the SNP markers according to the physical positions within the miscanthus reference genome, and a few badly ordered markers that were moved or removed according

to the recombination fractions between them, the final integrated map was made up of 2,602 markers and had a total length of 2,770 cM (Table 2; Fig. S2). The different markers on the map were as follows: 613 “Bri” markers of B3.7 type, 1,256 “Mal” markers of D1.10 type and 733 “Sil” markers of D2.15 type. In the 19 LGs, LG1 was the largest with a length of 217.3 cM and LG12a and LG12b were the shortest with a length of 49.3 and 32.6 cM, respectively. The average inter-marker distance was 1.06 cM when considering the 19 LGs, while LG4 showed the highest density with a mean interval of 0.72 cM between markers, which was in accordance with the highest number of 271 markers mapped on this LG (Table 2). For each LG, the heatmap showed the good quality of the marker order (Fig. S3).

Transgressive segregation was observed for biomass production and composition traits

The distribution of the BLUP and phenotypic means of the progeny were observed for biomass production and composition traits for each condition (Fig. 4; Fig. S4; Fig. S5a, b, c, d, e). These conditions were related to the climatic condition that occurred across the year, the age of the plants, and the location in which they were established. This resulted in a total of 13 conditions. Parental BLUP were reported for each trait and each condition. For both biomass production traits and composition traits, transgressive segregation was observed for each condition, except for the plant stem number (PSNb). In fact, the two parents of the population were chosen and crossed based on their highly contrasted stem number, as shown in Fig. 4. High genetic variability was observed for each of the biomass production traits: this was shown at age 3, by the percentage of the BLUP variation range compared to the corresponding phenotypic mean. These percentages ranged between 35% for the plant maximum height (HMax_cm) in Estrées-Mons to 186% for the aboveground biomass yield (ABM_tDMha) in Orléans. This variability tended to increase over the years (*i.e.* climatic conditions) and ages for most of the biomass production traits, based on the higher range of BLUP observed for each year or age than for the previous year or age.

Regarding biomass composition traits, their genetic variability was lower than the genetic variability observed for biomass production traits: the percentage of the BLUP range based on related phenotypic means varied between 5% for NDF_%DM in Estrées-Mons to 30% for ADL_%DM in Orléans. In contrast to the biomass production traits, this genetic variability did not increase across the years or ages.

Stable QTL were mainly detected across climatic conditions and ages for biomass production traits, while few stable QTL were detected for biomass composition traits

A total of 260 QTL was detected for biomass production traits (Table 3a; Table S1) and 283 QTL for biomass composition traits (Table 3b; Table S1). Overall, more QTL were detected in Estrées-Mons than in Orléans. The number of QTL was reported for each trait under each condition (Tables 3a and 3b). For each location, stable QTL (*i.e.* QTL that co-localized under at least two conditions) were highlighted across the years, ages and for both years and ages. An example is given with the solid red triangle in LG8 (37 cM) in Fig. 5: in Estrées-Mons, three stable QTL for aboveground biomass yield (ABM_tDMha) co-localized in 2016, 2017 and 2018, and three QTL also co-localized for age 2, age 3 and age 4. Six QTL thus co-localized for this trait and were either related to years and/or ages. Stable QTL were also

highlighted in both locations: this is the case for the plant stem number (PSNb) QTL in LG4 (125 cM), symbolized by an empty red triangle (Fig. 5). In this example, two QTL co-localized for age 2 in Estrées-Mons and Orléans. Considering these two QTL and a QTL detected in 2017 in Estrées-Mons, three QTL co-localized for climatic conditions and ages in both locations. Finally, another case occurred: for example, a QTL detected in 2017 in Estrées-Mons and in 2018 in Orléans, was considered as a stable QTL for years in both locations.

For biomass production traits considered in each location, the average proportion of stable QTL detected according to different years or different ages was similar, and around 30% (Table 3a). These proportions were mainly consistent between traits and conditions. The stable QTL were not necessarily the QTL that displayed the highest percentage of genotypic variation accounted for (R^2), or the most significant QTL. In each location, the proportion of stable QTL detected by considering years and ages together on the one hand, was higher than stable QTL that only considered years or ages separately, on the other hand. This proportion was 59% in Estrées-Mons and 39% in Orléans. Thus, QTL found in a given year (for example for the climatic condition in 2017 when G1 plants were 3 years old and G2 plants 2 years old), were also detected for a given age (*i.e.* age 3 for plants grown in 2017 for G1 and 2018 for G2). In both locations, a proportion of 2% of stable QTL were found across the successive climatic conditions as well as 15% across the different ages: as observed in each location considered separately, the consideration of the stability across successive climatic conditions and successive ages together, made it possible to slightly increase the proportion of stable QTL to up to 16%.

For the biomass composition traits evaluated in each location, the proportion of stable QTL was globally higher for the QTL related to the different climatic conditions, than for the QTL related to the different ages (Table 3b). However, these proportions were relatively low and never exceeded 15% on average, with some traits showing no stable QTL at all. As observed for biomass production traits, the consideration of successive climatic conditions and successive ages together, made it possible to increase the proportion of stable QTL for biomass composition traits to up to around 30% in each location. In both locations, a proportion of 10% of stable QTL were found for different years, 11% for different ages and 12% for different years and ages. For traits expressed either as %DM or %CW, a similar number of QTL and proportion of stable QTL were globally detected for the different conditions.

QTL clusters were identified for biomass production traits and biomass composition traits

Many trait QTL co-localized with other trait QTL under each of the 13 conditions. Therefore, QTL clusters were defined when at least three QTL from different traits co-localized. This led to the identification of 12 QTL clusters, that were either specific to biomass production traits, biomass composition traits, or both (Fig. 5 and Table 5a).

Five QTL clusters were identified for biomass production traits and were located in LG4, LG7 and LG18 (Fig. 5 and Table 5a). Four of these clusters were made up of QTL for canopy height (CH_cm), plant circumference (C50_cm) and aboveground biomass yield (ABM_tDMha). One of the clusters, located in

LG7, was made up of QTL for canopy height, plant maximum height (HMax_cm) and aboveground biomass yield. All these clusters for biomass production traits were identified in Estrées-Mons, in 2018 and at ages 1, 2, and 4. In LG18, a particularly stable cluster, based on the same component traits, was found in 2018, at age 2 and age 4. The reliability of the clusters was confirmed based on the large positive correlations between the BLUP of the traits that belonged to the clusters identified for each condition. These correlations were all significant and ranged from 0.65 to 0.96 (Fig. 6).

Regarding biomass composition traits, six QTL clusters were identified (Fig. 5 and Table 5a): three of them were located in LG4, LG5 as well as LG15 and were made up of QTL for cellulose, hemicellulose and lignin contents (expressed as %DM or %CW). Two other clusters were made up of QTL for Acid Detergent Fiber (ADF_%DM), cellulose (as %DM or %CW) and hemicelluloses (as %CW) and were located in LG13 and LG15. Finally, the last cluster was located in LG16 and was made up of QTL related to ADF_%DM, CL_%DM and ADL_%DM. As for biomass production traits, the QTL clusters were confirmed by the high, significant and positive or negative correlations between the biomass composition traits (Fig. 6): they ranged from -0.55 to -0.96 for negative values and from 0.48 to 0.98 for positive values (except for the correlations between cellulose and lignin, which ranged from 0.37 to 0.47). Stable QTL clusters were found in LG15, for QTL detected in 2017 and at age 2 in Orléans. The other QTL clusters were found in 2017, at age 2 and age 3 in Estrées-Mons and in 2018 in Orléans.

Interestingly, a QTL cluster made up of both biomass production and composition traits was found in LG15, for the year 2018 (*i.e.* climatic condition) in Orléans (Fig. 5). These QTL were related to canopy height (CH_cm), Neutral Detergent Fiber (NDF_%DM) and Hemicelluloses (HEM_%CW). This was consistent with the significant correlations between the corresponding BLUP: 0.44 between CH_cm and NDF_%DM and -0.55 between CH_cm and HEM_%CW (Fig. 6).

QTL effects of biomass production and composition traits were found to be stronger in Orléans than in Estrées-Mons

According to the QTL found for each of the biomass production and composition traits, the minimum and maximum additive effects of the two parents and the minimum and maximum dominance effects were reported for each location in Table 4, as well as the ranges of R^2 and LOD threshold values. The proportion of significant additive and dominance effects found based on all QTL corresponding to each trait was also reported.

The QTL were detected according to an LOD threshold that ranged from 5.0 to 11.5 on average. For biomass production traits, the highest percentage of genotypic variation explained by the QTL (R^2) of each trait ranged from 15.9–19.5% in Estrées-Mons and from 19.6–24.5% in Orléans. The QTL found in Orléans thus tended to explain more genotypic variation than the QTL found in Estrées-Mons: this was consistent with the values of additive and dominance effects, which were often higher in Orléans than in Estrées-Mons. For example, a maximal dominance effect of 8.2 cm was highlighted for total plant height in Orléans, compared to a BLUP of 5.3 cm in Estrées-Mons. Interestingly, most of the significant QTL

effects found for each production trait in both locations were either due to dominance effects or additive effects of the Silberspinne parent. This was notably observed for plant circumference measured at 50 cm (C50_cm), for which 43% of the significant effects originated from the Silberspinne allelic effect in Estrées-Mons (*i.e.* 23% from the Malepartus allelic effect and 34% from the dominance effect) and 46% originated from the dominance effect in Orléans.

Concerning biomass composition traits, the highest percentage of genotypic variation explained by the QTL of each trait ranged from 12–19.3% in Estrées-Mons and from 17.2–29.4% in Orléans. As for biomass production traits, the QTL found in Orléans explained a higher percentage of the genotypic variation than the QTL found in Estrées-Mons: this was also consistent with the higher allelic and dominance effects often observed in Orléans than in Estrées-Mons. As an example, the maximum dominance effect for cellulose (%CW) was 0.41% in Orléans compared to 0.17% in Estrées-Mons. In contrast to the proportion of significant effects observed for biomass production traits, the maximum proportion of significant effects observed for biomass composition traits were either due to the Malepartus allelic effect, Silberspinne allelic effect or dominance effect, and that was the case in both locations. This can be shown for hemicellulose content (%DM) in Orléans, with 41% of the significant effects originating from the dominance effect. Concerning cellulose content (%CW) in Estrées-Mons, 37% of the significant effects originated from the Silberspinne allelic effect. Regarding lignin content (%CW), 38% of the significant effects originated from the Malepartus allelic effect.

Some *M. sinensis* genes within the QTL clusters were orthologous to sorghum and maize cell-wall-related genes

A total of 809 and 494 *M. sinensis* genes within the QTL clusters were found to be orthologous to genes in sorghum and maize respectively. Finally, the QTL clusters contained 62 of these *M. sinensis* genes that were orthologous to cell-wall-related genes in sorghum and maize (Table 5b), knowing that the QTL clusters were either related to miscanthus biomass production or composition traits (Tables 5a and 5b). The QTL cluster located in LG5 did not contain any *M. sinensis* genes that were orthologous to cell-wall-related genes in sorghum and maize. Regarding the other clusters, some underlying genes were identified for different climatic conditions and different ages (Tables 5a and 5b). The 62 *M. sinensis* genes belonged to three main categories: 15 genes (24%) coded for enzymes involved in the polysaccharides biosynthesis, 11 genes (18%) coded for enzymes involved in the phenylpropanoid pathway that is required for the biosynthesis of lignin, and 5 genes (8%) coded for cell-wall proteins. Finally, 17 transcription factors (27%) were also identified based on the miscanthus literature review (Table 5b).

Discussion

The stability of the QTL detected for biomass production and composition traits was investigated based on 13 different conditions related to the staggered-start design established in each of the two locations. The QTL of both types of traits were found to be more stable for successive climatic conditions and ages considered together, compared to successive climatic conditions or successive ages considered

separately. The evaluation of each climatic condition and each age was made possible based on the staggered-start design: usually, other types of designs such as 'single-start' designs lead to the evaluation of plants for a given year, in which the related age and climatic condition cannot be partitioned. According to our design, the biomass production traits appeared to be more stable than the biomass composition traits across the conditions evaluated. However, the stability of the QTL across both contrasted locations was found to be relatively low. The QTL clusters representing co-localizations of QTL for biomass production and/or composition traits were identified across 13 different conditions. The corresponding intervals were screened for the underlying genes that correspond to orthologous cell-wall-related genes known in sorghum and maize.

Three main points will be discussed in this section: (1) the stability of biomass production and composition traits highlighted across the ages and the climatic conditions of the successive years studied, based on the staggered-start design established in each location; (2) the clusters of QTL for biomass production and composition traits that are consistent with the moderate to high genetic correlations highlighted between these traits, and (3) the different orthologous cell-wall-related genes that are known in sorghum and maize, two relatives of miscanthus, and that were found at the flanking markers, support the interval positions of our QTL clusters.

Stable QTL of biomass production and composition traits were highlighted across climatic conditions and ages based on the staggered-start design of each location, while stable QTL were detected in both locations

In each location, stable QTL that corresponded to the QTL detected for a given trait that co-localized under at least two conditions across different climatic conditions and/or across different ages, were identified for biomass production and composition traits. The assessment of QTL stability was an important objective of the study, which is why the staggered-start design was analyzed according to two different linear mixed models, related to each climatic condition and each age. It also made it possible to consider all the genotypes of the population in each location. Regarding biomass production traits, the different climatic conditions and ages considered together in each location made it possible to highlight 59% and 39% of stable QTL in Estrées-Mons and Orléans, respectively. These results can be compared to those reported by Gifford et al. [37] and Dong et al. [39], in which each of the years studied in their experimental designs were not partitioned into age and climatic condition effects. Gifford et al. [37] studied 13 biomass production traits in a *M. sinensis* population over two successive years, among which they identified 61% of stable QTL: 22 QTL re-discovered in 2012 out of 36 QTL detected in 2011. Dong et al. [39] established three interconnected miscanthus populations and carried out four different QTL analysis methods, either related to CIM or stepwise analyses, which led to the detection of 288, 264, 133 and 109 QTL for 14 biomass production traits across two years. In 2013, they re-identified from 48 to 56% of the QTL that had already been detected in 2012. When climatic conditions and ages were considered separately in each location of our study, around 30% of stable QTL were found either over the years or across the ages, regardless of the location. Accordingly, these lower proportions result from the

partition of each year studied into the corresponding climatic condition on the one hand, and the age on the other.

Regarding biomass composition traits, a higher proportion of stable QTL was also found across the climatic conditions and ages when they were considered together rather than separately. However, these different proportions ranged from 3–29%, which was relatively low compared to biomass production traits. Van der Weijde et al. [40] studied a *M. sinensis* population for traits related to biomass composition and conversion efficiency: they found 23% of stable QTL in two successive years: 20 out of 86 QTL were found in 2013 and 2014. These proportions confirm that, in miscanthus, the QTL of biomass composition traits seem to be less stable across different years (and ages) than those of biomass production traits.

In each location studied, the proportion of stable QTL for both types of traits across the climatic conditions and/or ages was rather low: this could be expected, as it has been previously shown that biomass production and composition traits can be affected by the variability related to plant age and environmental factors, such as the related climatic conditions that occur each year [73, 74]. However, these stable QTL mapped across the climatic conditions and/or ages would lead to the identification of relevant targets for Marker-Assisted Selection (MAS) programs. Among these, an interesting example was highlighted in LG8 (37 cM) for plant circumference (C50_cm) and aboveground biomass yield (ABM_tDMha), symbolized with a solid red triangle in Fig. 5: these stable QTL are interesting in terms of their stability over different ages and climatic conditions of successive years in Estrées-Mons, especially as they are stable for age 2, age 3 and age 4. It means that the future genetic material could be screened at a young age, in order to select individuals that show interesting alleles according to this QTL. It could thus speed up miscanthus breeding when based on an early selection of such individuals.

The proportion of stable QTL depends on environmental conditions, plant age and the genetic material considered, which is specific to each miscanthus study [37, 39, 40]. However, these last studies did not detect QTL for more than three years after establishment, and did not partition the year into age and climatic condition effects. Segura et al. [41] used a staggered-start design and carried out QTL mapping in order to dissect the apple tree architecture into genetic, ontogenetic and environmental effects, which made it possible to determine the genetic determinism of related traits, with regard to tree ontogeny and climatic conditions. To our knowledge, the present study uses a staggered-start design in miscanthus for the first time, in order to detect stable QTL related to different climatic conditions and/or different ages. Moreover, a staggered-start design was established in each of the two contrasted locations, which led to the detection of QTL across locations as well, as had never been done before in miscanthus.

Finally, when considering the different years (*i.e.* climatic conditions) and ages together across locations, maximum proportions of 12% and 16% of stable QTL were highlighted for biomass production and composition traits, respectively. These different stable QTL are interesting for miscanthus breeding programs, as they express themselves for different climatic conditions, ages and locations. As for the stable QTL identified for different climatic conditions and/or different ages in each location, these QTL could also be relevant targets for MAS programs: the example in LG4 (125 cM) is interesting (symbolized

with an empty triangle in Fig. 5), as a QTL related to plant stem number (PSNb) is stable across the two locations at age 2, meaning that it expresses itself in two contrasted locations for the same age. The fact that the proportion of these stable QTL was not higher than what had been found elsewhere, indicates that other environmental effects interact with the genetic determinism of biomass production and composition traits across locations. The different climatic and soil conditions could explain that, as the staggered-start design was established in a deep loam soil in Estrées-Mons, while the staggered-start design in Orléans was established in a sandy soil. In addition, the climate in Estrées-Mons is more influenced by the ocean than in Orléans: the differences in climatic conditions between both locations are presented according to different periods of the plant cycle in Raverdy et al. (submitted to BioEnergy Research). Moreover, 53 additional progenies were grown in Estrées-Mons compared to Orléans (Fig. 1): the genetic variability was therefore not totally similar between the two locations. This can also be a reason why the genotypic variations (R^2) explained by the QTL found in Orléans were mainly higher than those explained by the QTL found in Estrées-Mons. Moreover, significant genotype \times location interactions could explain variable proportions of stable QTL across locations. In a study comparing different miscanthus species across five different locations in Europe, Clifton-Brown et al. [75] and Lewandowski et al. [76] indeed highlighted genotype \times location interactions for biomass production and biomass composition traits, respectively. The establishment effect can also impact QTL detection power, as miscanthus is mature from around two to three years [19] or five years after establishment [77]: a substantial number of QTL was found from young to old plants in our study, which suggests that the effect due to the establishment may be limited within the location (*i.e.* in our study, age 1 was defined in 2015 for G1 plants established in 2014 and in 2016 for G2 plants established in 2015). However, the different establishment conditions between locations could also influence the proportion of stable QTL found across locations. Tejera et al. [44] used a staggered-start design and showed that the *M. \times giganteus* yield response to fertilization was influenced by establishment conditions in each location but not by the plant age.

In this study, each staggered-start design makes it possible to highlight a higher proportion of stable QTL for a range of climatic conditions and ages considered together rather than separately. However, the stability of QTL under these conditions is higher for biomass production traits than for biomass composition traits, studied together for the first time in a miscanthus mapping population. Across locations, some stable QTL were stable despite different environmental conditions such as climate, soil and establishment effects. This brings new insights into miscanthus breeding, as stable QTL are needed from different genetic material evaluated across different ages and climatic conditions: the comparison of stable QTL between studies would lead to the identification of the most significant genomic regions associated with biomass production and composition traits.

Clusters of QTL for biomass production and composition traits were consistent with the moderate to high genetic correlations highlighted between these traits

The QTL clusters identified for biomass production and composition traits were in agreement with the moderate to high genetic correlations highlighted between the traits. The QTL clusters related to biomass

production traits were identified in LG4, LG7 and LG18. They were made up of QTL that overlapped at similar positions, for traits such as canopy height, total plant height, plant circumference and aboveground biomass yield. The corresponding significant and moderate to high genetic correlations suggest that QTL overlapping is not random. Moreover, the stability of the QTL cluster is shown in LG18, as QTL were detected in 2018 and at ages 2 and 4 in Estrées-Mons. This is possible based on each staggered-start design evaluated over five years in two locations. Gifford et al. [37] identified QTL clusters in LG3 and LG6, that were re-identified in two subsequent years and that were consistent with high genetic and phenotypic correlations as well. These clusters were made up of QTL related to the plant circumference, stem diameter, plant stem number, aboveground biomass yield or characteristics of the leaves such as leaf width, length and area. These QTL identified for leaf-related traits are interesting: as canopy height refers to the height of the different leaves of the plant that contributes to yield-building, the QTL identified for canopy height in our study can in fact be related to different phenotypic characteristics of the leaves in our population. However, none of their different QTL clusters were common to our QTL clusters. Dong et al. [39] highlighted different QTL clusters found in their three miscanthus interconnected populations: these clusters were related to plant height, plant circumference, stem characteristics such as its volume and density, and the aboveground biomass yield. They were identified in various LG according to the population considered, and were in agreement with the moderate to high phenotypic and genetic correlations between these traits. For one of their populations originating from a cross between a *M. sinensis* and a *M. sacchariflorus* cultivar, they identified QTL clusters in LG4 and LG7: these LGs were common to the LGs in which we found QTL clusters for the same type of traits related to plant height, plant circumference and aboveground biomass yield. However, an investigation of the QTL cluster positions found in their study would be based on the alignment with the *M. sinensis* reference genome in order to confirm if the same genomic regions are involved.

Regarding biomass composition traits, we identified QTL clusters in LG4, LG5, LG13, LG15 and LG16, which were also in agreement with the significant moderate to high correlations between these traits. The stability of the clusters is also highlighted because two clusters were found in 2017 and for age 2 in Orléans. This is made possible based on the staggered-start design as well. They were located in LG15 and made up of traits related to cellulose, hemicelluloses, lignin and ADF contents. The co-localization of QTL related to ADF content with those related to cellulose and lignin content is not surprising, as ADF content represents the sum of cellulose and lignin contents [65]. Van der Weijde et al. [40] identified a major QTL cluster for traits related to conversion efficiency and composition traits: this cluster was located on chromosome 6 according to the *Sorghum bicolor* reference genome that was used in the construction of their two parental miscanthus genetic maps. In miscanthus, the corresponding chromosomes are chromosomes 11 and 12, as the miscanthus genome has been shown to be the result of chromosomal duplication and fusion based on the sorghum genome [36, 38, 45]. However, none of our QTL clusters are common with their QTL clusters, because we did not identify QTL clusters in LG11 and LG12.

Our study was conducted by considering biomass production and composition traits together: this made it possible to identify a QTL cluster in LG15, which was made up of both biomass production and

composition traits. The corresponding traits were canopy height, NDF (%DM) and hemicellulose content: the moderate and significant correlations of canopy height with these composition traits (respectively 0.44 and - 0.55) tend to validate the existence of this cluster. However, further analysis according to the genes underlying this cluster is necessary to confirm this assumption.

The QTL clusters identified for biomass production and composition traits could be explained by different genetic factors, such as the pleiotropic effects of the genes underlying these QTL or linked genes. Sometimes, these clusters can originate from genomic regions with segregation distortion, but this may not be possible in our study as we carefully filtered the distorted markers for the construction of our integrated genetic map. The staggered-start design led to the identification of QTL clusters located in LG4, LG5, LG7, LG13, LG15, LG16, LG18 for a range of climatic conditions and ages that consider biomass production and composition traits together, for the first time in miscanthus.

Orthologous cell-wall-related genes previously known in sorghum and maize made possible the identification of putative cell-wall-related genes in *M. sinensis*

Some of the 62 *M. sinensis* genes that were identified in the QTL clusters based on the orthologous cell-wall-related genes known in sorghum and maize, belong to interesting gene families. Twelve genes among the 62 genes belong to families that were previously found to contain genes involved in the secondary cell-wall (SCW) biosynthesis of miscanthus [78, 79] and are highlighted in bold in Table 5b. It can be noted that some genes are repeated in Table 5b as they belong to different QTL clusters. Hu et al. [78] carried out a transcriptome analysis of genes involved in secondary cell-wall biosynthesis in developing internodes of *M. lutarioriparius*: they highlighted different gene members in specific gene families. These gene families were 4CL, CCR and LAC that contain genes involved in the lignin biosynthesis based on the phenylpropanoid pathway [80]. They also identified the CSL and GT gene families that are involved in the biosynthesis of cellulose and hemicellulose components in plants. Finally, three other gene families in common with Hu et al. [78] were identified: the FLA gene family, for which genes are involved in cell-wall modification and assembly; the TFNAC and TFWRKY families that contain transcriptional factors for the regulation of secondary cell-wall development. In addition, Zeng et al. [79] identified major genes involved in the monolignol biosynthesis pathway in *M. x giganteus* based on genetic and transcriptional analyses. Just like the phenylpropanoid pathway, this monolignol pathway leads to the synthesis of monolignols that are involved in lignin biosynthesis [80]. Some of the major genes identified by Zeng et al. [79] belong to gene families in which we also found genes: the 4CL and CCR families that were also reported by Hu et al. [78], as well as the HCT family.

Based on these different comparisons, we can hypothesize that twelve *M. sinensis* genes out of the 62 genes previously identified are involved in secondary cell-wall development. This hypothesis can be supported by the fact that these genes were mainly found in the QTL clusters composed of *M. sinensis* biomass composition traits, especially for the clusters located in LG4, LG13 and LG15. Cluster 2 in LG4 and clusters 1 and 3 in LG15 were particularly interesting, as the R^2 of the related QTL mainly ranged from 11,3 to 29.4% (Table 5a).

Conclusion

In this study, an integrated genetic map of 2,770 cM was constructed based on 2,602 SNP markers distributed across 19 LGs, and was aligned with the released *M. sinensis* reference genome. This integrated genetic map, which was highly saturated, made it possible to identify 260 and 283 QTL related to biomass production and composition traits, respectively. The staggered-start design established in each of the two contrasted French locations, led to the detection of QTL that were stable across different climatic conditions and different ages. For both types of traits, a higher stability of the QTL was found when the climatic conditions were considered together with the different ages on the one hand, than when they were considered separately on the other. These differences were highlighted based on the partition of the year effect into age and climatic condition effects. The stability of the QTL across locations was found to be lower than the stability according to the conditions that were assessed within each location. However, the most stable QTL identified across both locations and across different climatic conditions and different ages would be interesting for miscanthus breeders, as they are stable regardless of the condition assessed in our experiment. They are thus important resources in order to carry out future Marker-Assisted Selection programs. This would be true especially for the QTL which were found stable at age 3 and age 4, as they could be interesting for screening young plants without the need to wait for their mature age. This would be more suited to biomass production traits, as the biomass composition traits were found to be less stable across the conditions. Regarding the unstable QTL that were specific to a given climatic condition or a given age, they may be explained by the existence of QTL that correspond to the genotype \times age and genotype \times climatic condition interaction effects. These effects were specifically assessed in the two models carried out for the analysis of the staggered-start designs, but their corresponding mapping has not been carried out yet, and their future detection would be desirable.

Interesting clusters of QTL were then identified for biomass production and composition traits under different conditions: this means that linked genes or pleiotropic effects from the genes underlying these QTL would make it possible to jointly improve these different traits. Moreover, these QTL clusters contained 62 *M. sinensis* genes that were orthologous to cell-wall-related genes in sorghum and maize. Twelve of these genes were identified as putatively involved in secondary cell-wall biosynthesis. In summary, all these QTL clusters which correspond to different traits or stability types and their underlying candidate genes constitute targets of interest for miscanthus breeders, in order to evaluate and create new miscanthus cultivars that would be adapted to different environments, with a high biomass yield and a composition suited to bioenergy, biomaterials or animal bedding.

Declarations

Author contributions

Maryse Brancourt-Hulmel carried out the conception and the design of the study. Kristelle Lourgant, Emilie Mignot, Stéphanie Arnoult, Guillaume Bodineau and Raphaël Raverdy prepared the material and

carried out phenotyping data management and collection. Yves Griveau realized NIRS data collection and predictions. Raphaël Raverdy, Maryse Brancourt-Hulmel, Kristelle Lourgant and Cristiane H. Taniguti contributed to the analysis of the data. Raphaël Raverdy wrote the manuscript and all authors commented on the manuscript.

Acknowledgments

The authors wish to acknowledge each member of the staff who worked for the establishment and phenotyping of each trial in INRAE experimental stations: the GCIE–Picardie team in Estrées-Mons, the GBFO team in Orléans and the IJPB team in Versailles. The authors are grateful to the teams of the INRAE Gentyane Genomic Services platform (Clermont-Ferrand, Puy-de-Dôme, France), the CIRAD Genotyping platform core facility (GPTR, Montpellier, Hérault, France) and the GenoToul GeT platform (Auzerville, Haute-Garonne, France) for DNA extraction, GBS data and sequencing, respectively. The authors also acknowledge David Pot and Matthieu Reymond for providing lists of cell-wall-related genes identified in sorghum and maize. Finally, they thank Rebecca James who edited the English of the manuscript.

References

1. Jones M, Walsh M (2001) *Miscanthus for Energy and Fibre*. James & James (Science Publishers) Ltd.
2. Heaton EA, Clifton-Brown J, Voigt TB et al (2004) *Miscanthus for renewable energy generation: European Union experience and projections for Illinois*. *Mitig Adapt Strateg Glob Chang* 9:433–451. <https://doi.org/10.1023/B:MITI.0000038848.94134.be>
3. Johnson M, Tucker N, Barnes S, Kirwan K (2005) Improvement of the impact performance of a starch based biopolymer via the incorporation of *Miscanthus giganteus* fibres. *Ind Crops Prod* 22:175–186. <https://doi.org/10.1016/j.indcrop.2004.08.004>
4. Girones J, Vo L, Arnoult S et al (2016) *Miscanthus stem fragment – Reinforced polypropylene composites: Development of an optimized preparation procedure at small scale and its validation for differentiating genotypes*. *Polym Test* 55:166–172. <https://doi.org/10.1016/j.polymertesting.2016.08.023>
5. Lewandowski I, Clifton-Brown J, Kiesel A et al (2018) *Miscanthus*. *Perenn Grasses Bioenergy Bioprod* 23:35–59. <https://doi.org/10.1016/b978-0-12-812900-5.00002-3>
6. Acikel H (2011) The use of *Miscanthus (Giganteus)* as a plant fiber in concrete production. *Sci Res Essays* 6:2660–2667. <https://doi.org/10.5897/SRE10.1139>
7. Anderson E, Arundale R, Maughan M et al (2011) Growth and agronomy of *Miscanthus × giganteus* for biomass production. *Biofuels* 2:167–183. <https://doi.org/10.4155/bfs.10.80>
8. Hodkinson TR, Klaas M, Jones MB et al (2015) *Miscanthus: A case study for the utilization of natural genetic variation*. *Plant Genet Resour Characterisation Util* 13:219–237.

<https://doi.org/10.1017/S147926211400094X>

9. Clifton-Brown JC, Neilson B, Lewandowski I, Jones MB (2000) The modelled productivity of *Miscanthus × giganteus* (GREEF et DEU) in Ireland. *Ind Crop Prod* 12:97–109
10. Heaton EA, Dohleman FG, Long SP (2008) Meeting US biofuel goals with less land: The potential of *Miscanthus*. *Glob Chang Biol* 14:2000–2014. <https://doi.org/10.1111/j.1365-2486.2008.01662.x>
11. Greef J, Deuter M (1993) Syntaxonomy of *Miscanthus x giganteus* Greed et Deu. *Angew Bot* 67:87–90
12. Linde-Laursen I (1993) Cytogenetic Analysis of *Miscanthus* ‘Giganteus’, an Interspecific Hybrid. *Hereditas* 119:297–300. <https://doi.org/10.1111/j.1601-5223.1993.00297.x>
13. Sun Q, Lin Q, Yi Z-L et al (2010) A taxonomic revision of *Miscanthus* s.l. (Poaceae) from China. *Bot J Linn Soc* 164:178–220. <https://doi.org/10.1111/j.1095-8339.2010.01082.x>
14. Sacks EJ, Juvik JA, Lin Q et al (2013) The Gene Pool of *Miscanthus* Species and Its Improvement. In: Paterson A (ed) *Genomics of the Saccharinae. Plant Genetics and Genomics: Crops and Models*, vol 11. Springer, New York
15. Clark LV, Brummer JE, Głowacka K et al (2014) A footprint of past climate change on the diversity and population structure of *Miscanthus sinensis*. *Ann Bot* 114:97–107. <https://doi.org/10.1093/aob/mcu084>
16. Clark LV, Ryan Stewart J, Nishiwaki A et al (2015) Genetic structure of *Miscanthus sinensis* and *Miscanthus sacchariflorus* in Japan indicates a gradient of bidirectional but asymmetric introgression. *J Exp Bot* 66:4213–4225. <https://doi.org/10.1093/jxb/eru511>
17. Anzoua KG, Suzuki K, Fujita S et al (2015) Evaluation of morphological traits, winter survival and biomass potential in wild Japanese *Miscanthus sinensis* Anders. populations in northern Japan. *Grassl Sci* 61:83–91. <https://doi.org/10.1111/grs.12085>
18. Głowacka K (2011) A review of the genetic study of the energy crop *Miscanthus*. *Biomass Bioenerg* 35:2445–2454. <https://doi.org/10.1016/j.biombioe.2011.01.041>
19. Clifton-brown JC, Chiang Y, Hodkinson TR (2008) *Miscanthus* genetic resources and breeding potential. In: Vermeris W (ed) *Genetic Improvement of Bioenergy Crops*. Springer Science, pp 273–290
20. Zub HW, Brancourt-Hulmel M (2010) Agronomic and physiological performances of different species of *Miscanthus*, a major energy crop. A review. *Agron Sustain Dev* 30:201–214. <https://doi.org/10.1051/agro/2009034>
21. Conson ARO, Taniguti CH, Amadeu RR et al (2018) High-Resolution Genetic Map and QTL Analysis of Growth-Related Traits of *Hevea brasiliensis* Cultivated Under Suboptimal Temperature and Humidity Conditions. *Front Plant Sci* 9:1–16. <https://doi.org/10.3389/fpls.2018.01255>
22. Balsalobre TWA, da Silva Pereira G, Margarido GRA et al (2017) GBS-based single dosage markers for linkage and QTL mapping allow gene mining for yield-related traits in sugarcane. *BMC Genom* 18:1–19. <https://doi.org/10.1186/s12864-016-3383-x>

23. Souza LM, Gazaffi R, Mantello CC et al (2013) QTL Mapping of Growth-Related Traits in a Full-Sib Family of Rubber Tree (*Hevea brasiliensis*) Evaluated in a Sub-Tropical Climate. *PLoS One* 8:. <https://doi.org/10.1371/journal.pone.0061238>
24. 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
25. Maliepaard C, Jansen J, Van Ooijen JW (1997) Linkage analysis in a full-sib family of an outbreeding plant species: Overview and consequences for applications. *Genet Res* 70:237–250. <https://doi.org/10.1017/S0016672397003005>
26. Wu R, Ma CX, Painter I, Zeng ZB (2002) Simultaneous Maximum Likelihood Estimation of Linkage and Linkage Phases in Outcrossing Species. *Theor Popul Biol* 61:349–363. <https://doi.org/10.1006/tpbi.2002.1577>
27. Wu R, Ma CX, Wu SS, Zeng ZB (2002) Linkage mapping of sex-specific differences. *Genet Res* 79:85–96. <https://doi.org/10.1017/S0016672301005389>
28. Margarido GRA, Souza AP, Garcia AAF (2007) OneMap: Software for genetic mapping in outcrossing species. *Hereditas* 144:78–79. <https://doi.org/10.1111/j.2007.0018-0661.02000.x>
29. Gazaffi R, Amadeu RR, Mollinari M et al (2020) fullsibQTL: an R package for QTL mapping in biparental populations of outcrossing species
30. Gazaffi R, Margarido GRA, Pastina MM et al (2014) A model for quantitative trait loci mapping, linkage phase, and segregation pattern estimation for a full-sib progeny. *Tree Genet Genomes* 10:791–801. <https://doi.org/10.1007/s11295-013-0664-2>
31. Atienza SG, Satovic Z, Petersen KK et al (2002) Preliminary genetic linkage map of *Miscanthus sinensis* with RAPD markers. *Theor Appl Genet* 105:946–952. <https://doi.org/10.1007/s00122-002-0956-7>
32. Atienza SG, Satovic Z, Petersen KK et al (2003) Identification of QTLs influencing agronomic traits in *Miscanthus sinensis* Anders. I. Total height, flag-leaf height and stem diameter. *Theor Appl Genet* 107:123–129. <https://doi.org/10.1007/s00122-003-1220-5>
33. Atienza SG, Satovic Z, Petersen KK et al (2003) Identification of QTLs associated with yield and its components in *Miscanthus sinensis* Anders. *Euphytica* 132:353–361. <https://doi.org/10.1023/A:1025041926259>
34. Atienza SG, Satovic Z, Petersen KK et al (2003) Identification of QTLs influencing combustion quality in *Miscanthus sinensis* Anders. II. Chlorine and potassium content. *Theor Appl Genet* 107:857–863. <https://doi.org/10.1007/s00122-003-1218-z>
35. Atienza SG, Satovic Z, Petersen KK et al (2003) Influencing combustion quality in *Miscanthus sinensis* Anders.: Identification of QTLs for calcium, phosphorus and sulphur content. *Plant Breed* 122:141–145. <https://doi.org/10.1046/j.1439-0523.2003.00826.x>
36. Ma XF, Jensen E, Alexandrov N et al (2012) High resolution genetic mapping by genome sequencing reveals genome duplication and tetraploid genetic structure of the diploid *Miscanthus sinensis*. *PLoS One* 7:. <https://doi.org/10.1371/journal.pone.0033821>

37. Gifford JM, Chae WB, Swaminathan K et al (2015) Mapping the genome of *Miscanthus sinensis* for QTL associated with biomass productivity. *GCB Bioenergy* 7:797–810. <https://doi.org/10.1111/gcbb.12201>
38. Swaminathan K, Chae WB, Mitros T et al (2012) A framework genetic map for *Miscanthus sinensis* from RNAseq-based markers shows recent tetraploidy. *BMC Genom* 13:142. <https://doi.org/10.1186/1471-2164-13-142>
39. Dong H, Liu S, Clark LV et al (2018) Genetic mapping of biomass yield in three interconnected *Miscanthus* populations. *GCB Bioenergy* 10:165–185. <https://doi.org/10.1111/gcbb.12472>
40. Van der Weijde T, Kamei CLA, Severing EI et al (2017) Genetic complexity of *Miscanthus* cell wall composition and biomass quality for biofuels. *BMC Genom* 18:1–15. <https://doi.org/10.1186/s12864-017-3802-7>
41. Segura V, Durel C, Costes E (2009) Dissecting apple tree architecture into genetic, ontogenetic and environmental effects : QTL mapping. *Tree Genet Genomes* 165–179. <https://doi.org/10.1007/s11295-008-0181-x>
42. Loughin TM (2006) Improved experimental design and analysis for long-term experiments. *Crop Sci* 46:2492–2502. <https://doi.org/10.2135/cropsci2006.04.0271>
43. Segura V, Cilas C, Segura V et al (2008) Dissecting apple tree architecture into genetic, ontogenetic and environmental effects: mixed linear modelling of repeated spatial and temporal measures. *New Phytol*. <https://doi.org/10.1111/j.1469-8137.2007.02374.x>
44. Tejera M, Boersma N, Vanlooche A et al (2019) Multi-year and Multi-site Establishment of the Perennial Biomass Crop *Miscanthus × giganteus* Using a Staggered Start Design to Elucidate N Response. *Bioenergy Res* 12:471–483. <https://doi.org/10.1007/s12155-019-09985-6>
45. Mitros T, Session AM, James BT et al (2020) Genome biology of the paleotetraploid perennial biomass crop *Miscanthus*. *Nat Commun* 1–11. <https://doi.org/10.1038/s41467-020-18923-6>
46. Dagnelie P (2012) Principes d'expérimentation: planification des expériences et analyse de leurs résultats, Les Presse
47. Elshire RJ, Glaubitz JC, Sun Q et al (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6:1–10. <https://doi.org/10.1371/journal.pone.0019379>
48. Cormier F, Lawac F, Maledon E et al (2019) A reference high-density genetic map of greater yam (*Dioscorea alata* L.). *Theor Appl Genet* 132:1733–1744. <https://doi.org/10.1007/s00122-019-03311-6>
49. Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
50. Glaubitz JC, Casstevens TM, Lu F et al (2014) TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. *PLoS One* 9:. <https://doi.org/10.1371/journal.pone.0090346>
51. Langmead B, Salzberg S (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 357–359. <https://doi.org/10.1038/nmeth.1923>

52. Garsmeur O, Droc G, Antonise R et al (2018) A mosaic monoploid reference sequence for the highly complex genome of sugarcane. *Nat Commun* 9:. <https://doi.org/10.1038/s41467-018-05051-5>
53. Margarido GRA, Mollinari M, Broman K et al (2020) onemap: Software for constructing genetic maps in experimental crosses: full-sib, RILs, F2 and backcrosses. R package version 2.2.0
54. Kosambi DD (1943) The estimation of map distances from recombination values. *Ann Eugen* 12:172–175. <https://doi.org/10.1111/j.1469-1809.1943.tb02321.x>
55. Mollinari M, Garcia AAF (2019) Linkage analysis and haplotype phasing in experimental autopolyploid populations with high ploidy level using hidden Markov models. *G3 Genes, Genomes, Genet* 9:3297–3314. <https://doi.org/10.1534/g3.119.400378>
56. Mollinari M, Olukolu BA, da Pereira GS et al (2020) Unraveling the hexaploid sweetpotato inheritance using ultra-dense multilocus mapping. *G3 Genes, Genomes Genet* 10:281–292. <https://doi.org/10.1101/689638>
57. Lander ES, Green P (1987) Construction of multilocus genetic linkage maps in humans. *Proc Natl Acad Sci U S A* 84:2363–2367. <https://doi.org/10.1073/pnas.84.8.2363>
58. Preedy KF, Hackett CA (2016) A rapid marker ordering approach for high-density genetic linkage maps in experimental autotetraploid populations using multidimensional scaling. *Theor Appl Genet* 129:2117–2132. <https://doi.org/10.1007/s00122-016-2761-8>
59. Preedy KF, Hackett CA, Boskamp B (2018) MDSMap: High density linkage maps using multi-dimensional scaling. 1–10
60. Hackett CA, Broadfoot LB (2003) Effects of genotyping errors, missing values and segregation distortion in molecular marker data on the construction of linkage maps. *Heredity* 90:33–38. <https://doi.org/10.1038/sj.hdy.6800173>
61. Cartwright DA, Troggio M, Velasco R, Gutin A (2007) Genetic mapping in the presence of genotyping errors. *Genetics* 176:2521–2527. <https://doi.org/10.1534/genetics.106.063982>
62. Bilton TP, Schofield MR, Black MA et al (2018) Accounting for Errors in Low Coverage High-Throughput Sequencing Data when Constructing Genetic Maps using Biparental Outcrossed Populations. *bioRxiv* 209:65–76. <https://doi.org/10.1101/249722>
63. Taniguti CH (2021) Building highly saturated genetic maps with OneMap 3.0. new approaches using workflows
64. Steinbach D, Alaux M, Amselem J et al (2013) GnpIS: An information system to integrate genetic and genomic data from plants and fungi. *Database* 2013:1–9. <https://doi.org/10.1093/database/bat058>
65. Van Soest PJ, Wine RH (1967) Use of Detergents in the Analysis of Fibrous Feeds. IV. Determination of Plant Cell-Wall Constituents. *J AOAC Int* 50:50–55. <https://doi.org/10.1093/jaoac/50.1.50>
66. Henderson CR (1984) *Applications of Linear Models in Animal Breeding*. Univ Guelph, Guelph
67. Muñoz F, Sanchez L (2019) breedR: Statistical Methods for Forest Genetic Resources Analysts. R package version 0.12-4

68. Costa e Silva J, Dutkowski GW, Gilmour AR (2001) Analysis of early tree height in forest genetic trials is enhanced by including a spatially correlated residual. *Can J For Res* 31:1887–1893.
<https://doi.org/10.1139/cjfr-31-11-1887>
69. Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Automat Contr* 19:716–723. <https://doi.org/10.1109/TAC.1974.1100705>
70. Chen L, Storey JD (2006) Relaxed significance criteria for linkage analysis. *Genetics* 173:2371–2381.
<https://doi.org/10.1534/genetics.105.052506>
71. Lander ES, Botstein S (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121:185. <https://doi.org/10.1093/genetics/121.1.185>
72. Wei T, Simko V (2017) R package “corrplot”: Visualization of a Correlation Matrix (version 0.84)
73. Allison GG, Morris C, Clifton-Brown J et al (2011) Genotypic variation in cell wall composition in a diverse set of 244 accessions of *Miscanthus*. *Biomass Bioenerg* 35:4740–4747.
<https://doi.org/10.1016/j.biombioe.2011.10.008>
74. Arnoult S, Brancourt-Hulmel M (2015) A Review on *Miscanthus* Biomass Production and Composition for Bioenergy Use: Genotypic and Environmental Variability and Implications for Breeding. *Bioenergy Res* 8:502–526. <https://doi.org/10.1007/s12155-014-9524-7>
75. Clifton-Brown JC, Lewandowski I, Andersson B et al (2001) Performance of 15 *Miscanthus* genotypes at five sites in Europe. *Agron J* 93:1013–1019
76. Lewandowski I, Andersson B, Basch G et al (2003) Environment and Harvest Time Affects the Combustion Qualities of *Miscanthus* Genotypes. *Agron J* 1274–1280
77. Lesur C, Jeuffroy MH, Makowski D et al (2013) Modeling long-term yield trends of *Miscanthus* × *giganteus* using experimental data from across Europe. *F Crop Res* 149:252–260.
<https://doi.org/10.1016/j.fcr.2013.05.004>
78. Hu R, Xu Y, Yu C et al (2017) Transcriptome analysis of genes involved in secondary cell wall biosynthesis in developing internodes of *Miscanthus lutarioriparius*. *Sci Rep* 7:1–16.
<https://doi.org/10.1038/s41598-017-08690-8>
79. Zeng X, Sheng J, Zhu F et al (2020) Genetic, transcriptional, and regulatory landscape of monoglucosyl biosynthesis pathway in *Miscanthus* × *giganteus*. *Biotechnol Biofuels* 13:1–14.
<https://doi.org/10.1186/s13068-020-01819-4>
80. Vanholme R, Demedts B, Morreel K et al (2010) Lignin biosynthesis and structure. *Plant Physiol* 153:895–905. <https://doi.org/10.1104/pp.110.155119>

Tables

Due to technical limitations, table 1 to 5 is only available as a download in the Supplemental Files section.

Figures

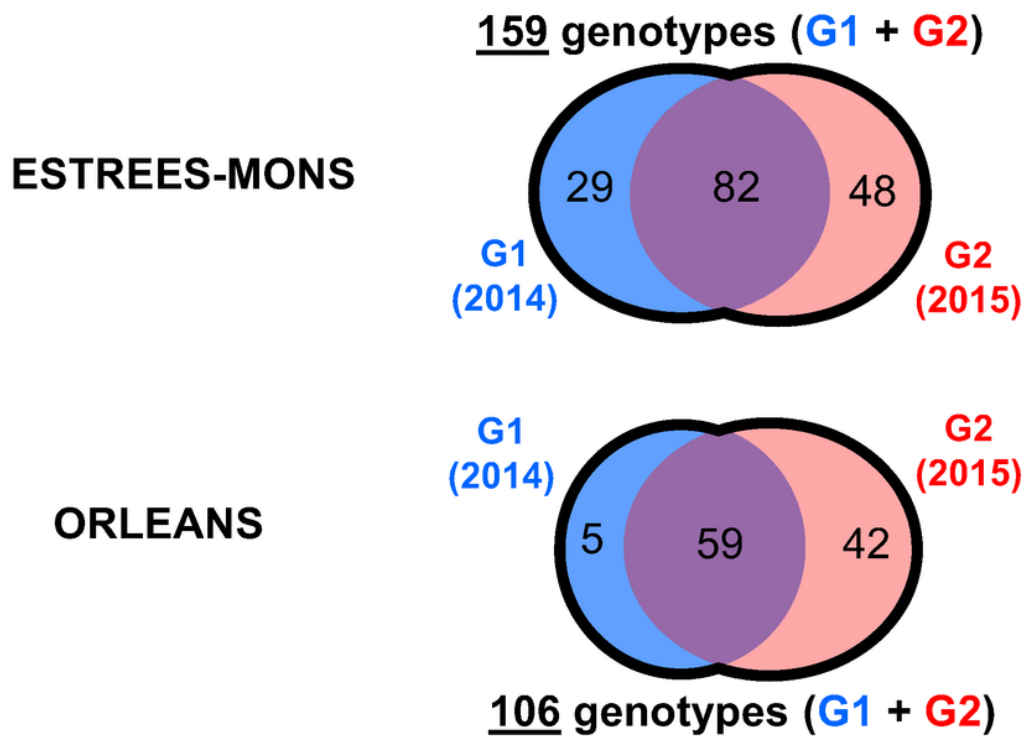


Figure 1

A staggered-start design was established in each location: Estrées-Mons and Orléans. The total number of *M. sinensis* genotypes (including the two parents) is indicated for each staggered-start design. The number of genotypes for each group (G1 and G2) and the genotypes common to both groups (at the intersection of blue and red circles) are also specified for each location.

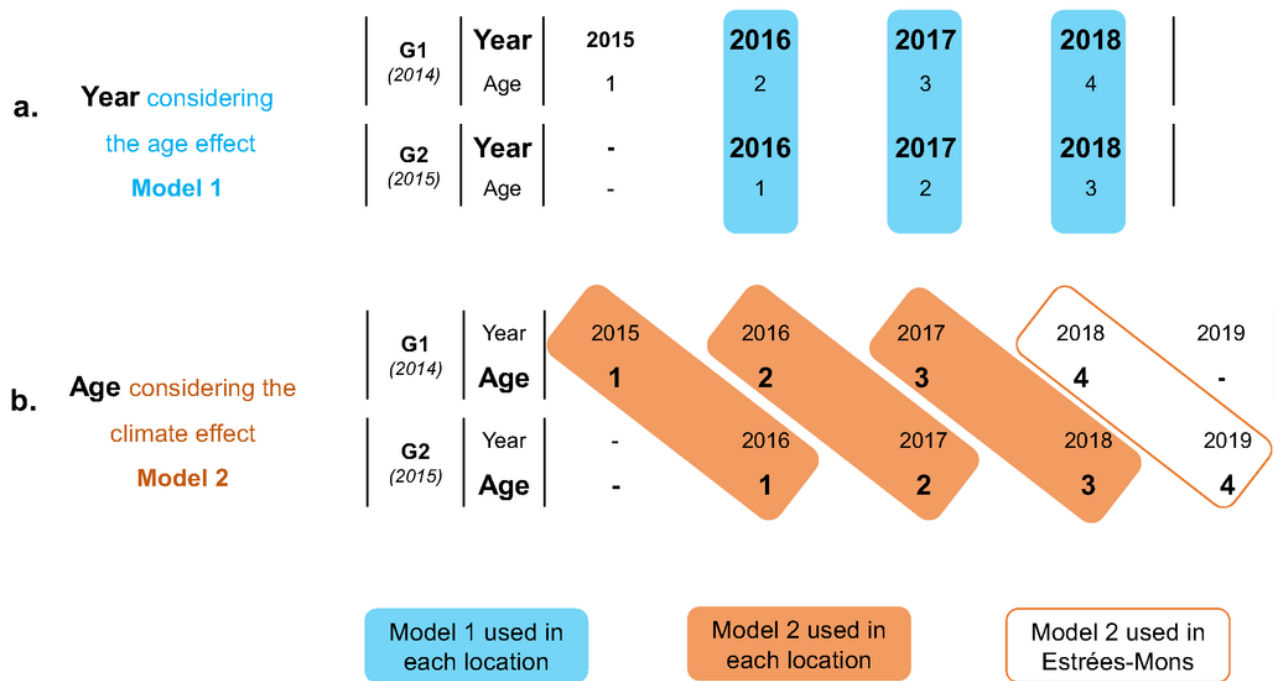


Figure 2

For each location considered separately, the corresponding staggered-start design was analyzed according to two statistical models: (a) the age effect model per growing season or year (i.e. climatic condition) for Model 1 and (b) the climatic condition effect model per age for Model 2. These models were based on two different subsets of the data. For example, case (a) was based on year 2016 considered for G1 and G2, with the age effect modeled according to 2-year-old genotypes in G1 and 1-year-old genotypes in G2. In this case, plants of different ages grew under the same climatic condition during a single year. While case (b) was based for example on genotypes with the same age (2-year-old), according to genotypes of G1 which grew in the year 2016 and genotypes of G2 which grew in the year 2017. For this case, plants of the same age grew under two different climatic conditions, related to each year considered for each group. The group year establishment is specified in brackets below each group name.

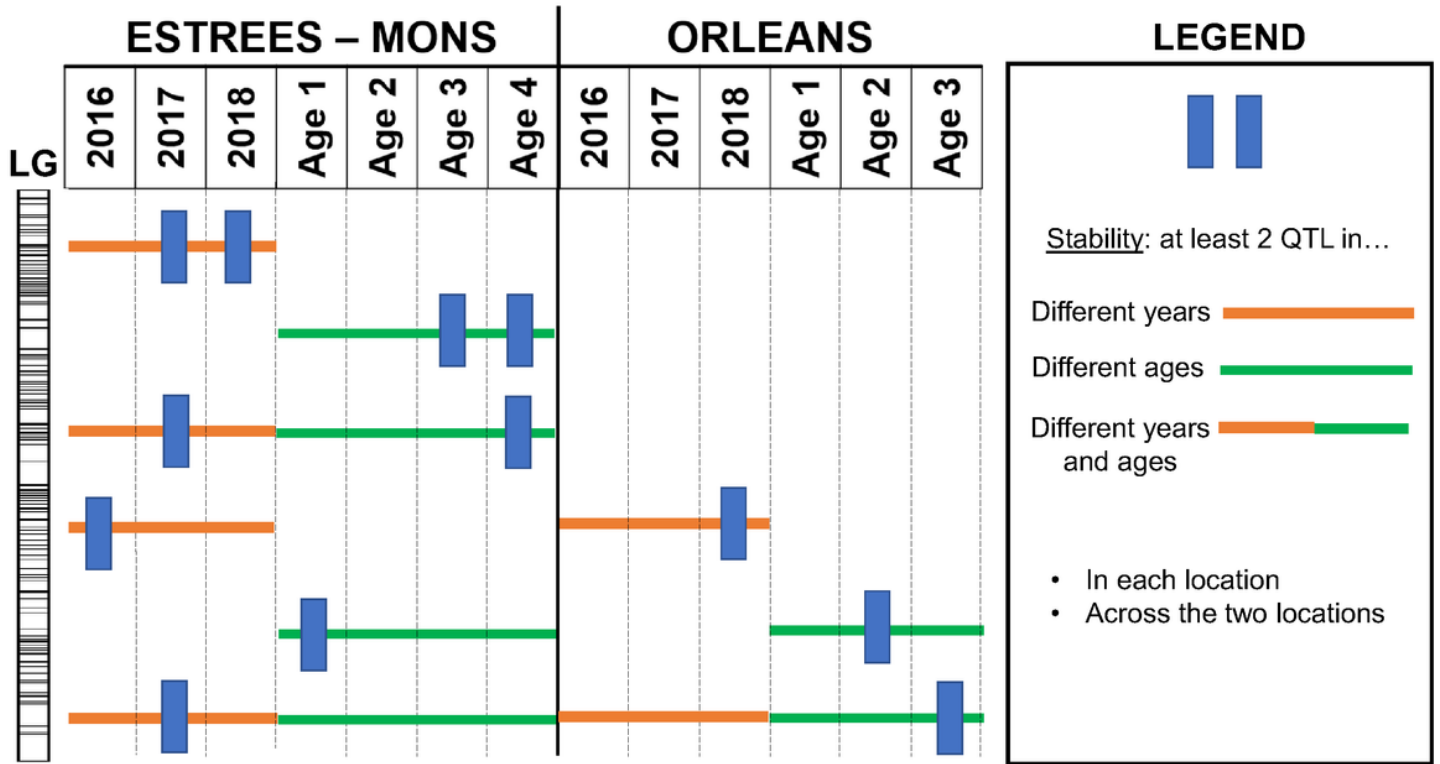


Figure 3

Illustration of the different types of QTL stability according to the 13 conditions studied.

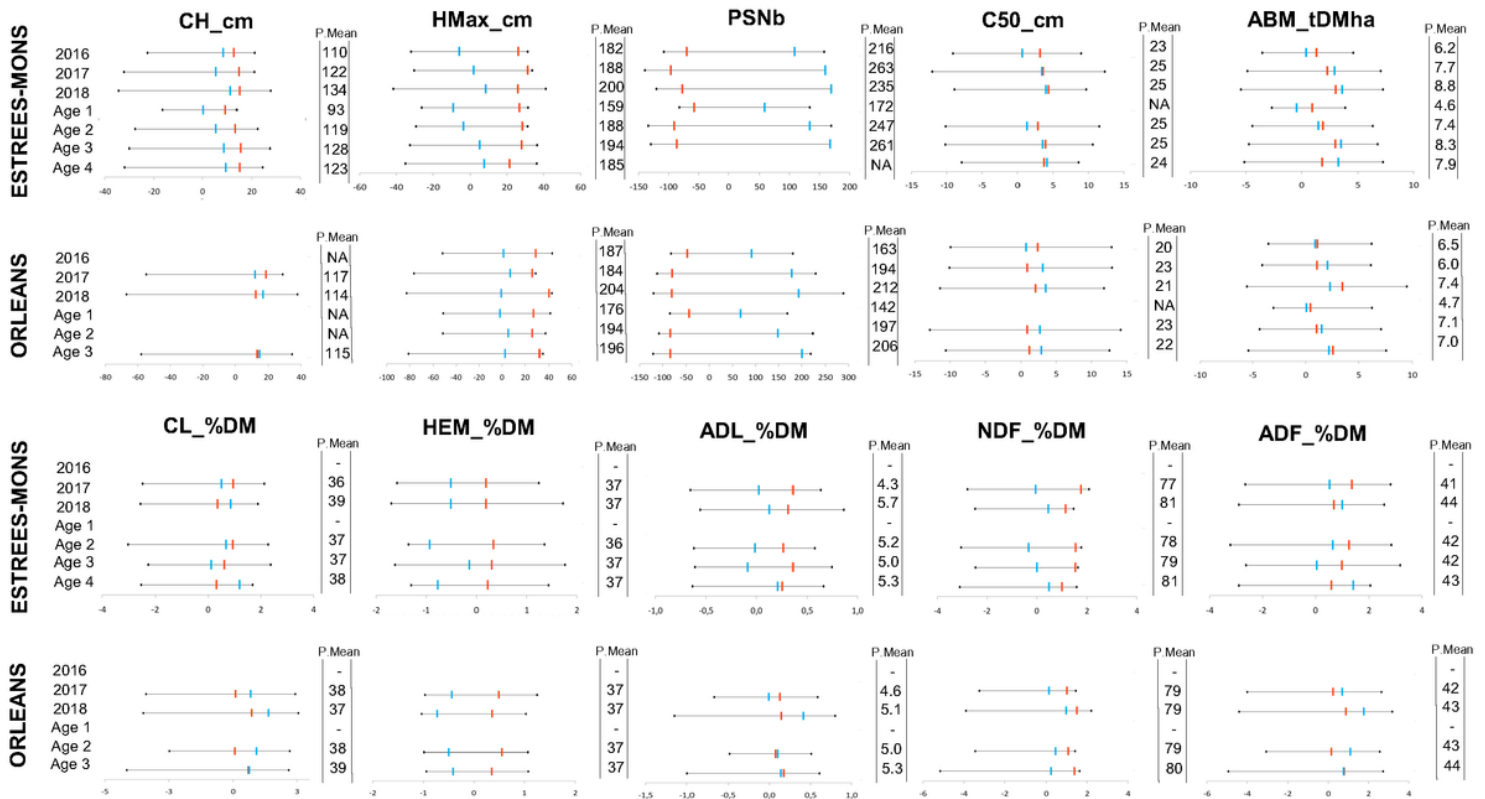


Figure 4

Summarized distributions of Best Linear Unbiased Predictors (BLUP) and the associated phenotypic means (P.Mean), for biomass production and composition traits for each condition. The red and blue lines represent the parental BLUP, which correspond to Malepartus and Silberspinne, respectively.



Figure 5

Representation of 8 LGs out of 19. The QTL were detected according to 13 conditions. A part of the stable QTL is illustrated, as well as QTL clusters. The QTL detected for a threshold based on a 10% significance level are marked with a “*” (see the Materials and Methods section for more details). The length of each LG is specified to the left of the LG in cM.

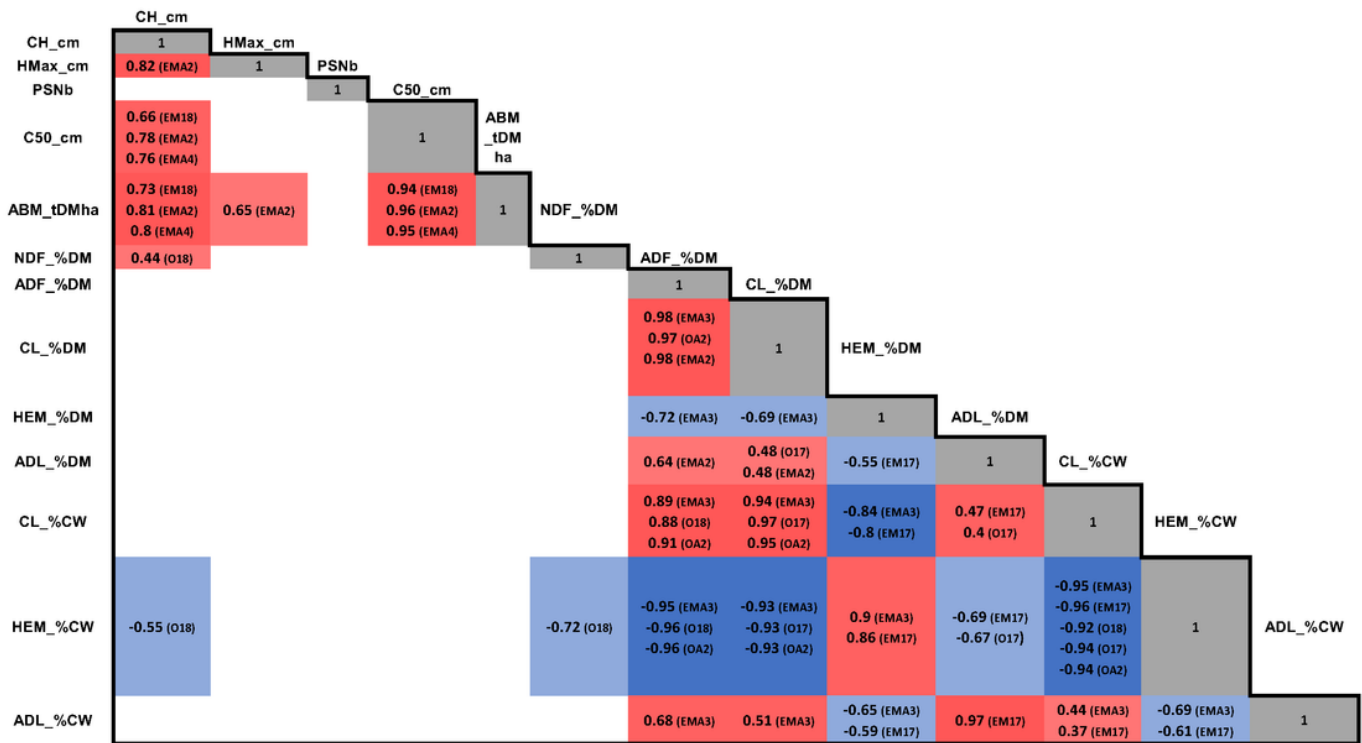


Figure 6

Significant correlations between BLUP that corresponded to QTL traits within the clusters, among all significant correlations. The significance level was assessed according to a p-value < 0.05. The blanks represent correlations that were not shown because the corresponding traits were not identified in the clusters. The conditions under which the QTL clusters were identified, are described as EM17: Estrées-Mons in 2017, EM18: Estrées-Mons in 2018, EMA2: Estrées-Mons at age 2, EMA3: Estrées-Mons at age 3, EMA4: Estrées-Mons at age 4, O17: Orléans in 2017, O18: Orléans in 2018 and OA2: Orléans at age 2. See the Materials and Methods section for details of trait names.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Table1.pdf](#)
- [Table2.pdf](#)
- [Table3.pdf](#)

- Table4.pdf
- Table5.pdf
- Fig.S1.pdf
- Fig.S2.pdf
- Fig.S3.pdf
- Fig.S4.pdf
- Fig.S5a.pdf
- Fig.S5b.pdf
- Fig.S5c.pdf
- Fig.S5d.pdf
- Fig.S5e.pdf
- TableS1.pdf