In silico QTL mapping in an oil palm breeding program reveals a quantitative and complex genetic resistance to Ganoderma boninense

Aurélie Daval, Virgine Pomiès, Sandrine Le Squin, Marie Denis, Virginie Riou, Frédéric Breton, - Nopariansyah, Bink Marco, Benoît Cochard, Florence Jacob, et al.

Title Page

Title

In silico QTL mapping in an oil palm breeding program reveals a quantitative and complex genetic resistance to *Ganoderma boninense*

Authors

Aurélie Daval, Virgine Pomiès, Sandrine le Squin, Marie Denis, Virginie Riou, Frederic Breton, Nopariansyah, Marco Bink, Benoît Cochard, Florence Jacob, Norbert Billotte and Sébastien Tisné

Author information

Aurélie Daval : CIRAD, UMR AGAP, F-34398 Montpellier, France ; AGAP, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France ; ORCID ID: 0000-0002-2613-0562

Virgine Pomiès : CIRAD, UMR AGAP, F-34398 Montpellier, France ; AGAP, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France ; ORCID ID: 0000-0002-5481-5120

Sandrine le Squin : PalmElit SAS, Montferrier-sur-Lez, France

Marie Denis : CIRAD, UMR AGAP, F-34398 Montpellier, France ; AGAP, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France ; ORCID ID : 0000-0002-1693-9894

Virginie Riou : CIRAD, UMR AGAP, F-34398 Montpellier, France ; AGAP, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France

Frederic Breton : CIRAD, UMR AGAP, F-34398 Montpellier, France ; AGAP, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France ; ORCID ID: 0000-0002-6853-2623

Nopariansyah: P.T SOCFINDO, Jl. Yos Sudarso, Medan, Sumatera Utara 20115, Indonesia

Marco Bink: Biometris, Wageningen UR, PO Box 16, 6700 AA Wageningen, The Netherlands; Current address: Research & Technology Center, Hendrix Genetics, Boxmeer, The Netherlands ; ORCID ID: 0000-0002-1278-2092

Benoît Cochard : PalmElit SAS, Montferrier-sur-Lez, France

Florence Jacob : PalmElit SAS, Montferrier-sur-Lez, France ; ORCID ID : 0000-0002-0454-1037

Norbert Billotte : CIRAD, UMR AGAP, F-34398 Montpellier, France ; AGAP, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France ; ORCID ID : 0000-0002-0438-0966

Sébastien Tisné (Corresponding author): CIRAD, UMR AGAP, F-34398 Montpellier, France ; AGAP, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France; E-mail address: sebastien.tisne@cirad.fr; ORCID ID: 0000-0001-9838-3975

Abstract

Basal stem rot caused by *Ganoderma boninense* is the major threat to oil palm cultivation in South-East Asia, which accounts for 80% of palm oil production worldwide, and this disease is increasing in Africa. The use of resistant planting material as part of an integrated pest management of this disease is one sustainable solution. However, breeding for *Ganoderma* resistance requires long-term and costly research, which could greatly benefit from marker assisted selection (MAS). In this study, we evaluated the effectiveness of an *in silico* genetic mapping approach that took advantage of extensive data recorded in an ongoing breeding program. A pedigree-based QTL mapping approach applied to more than 10 years' worth of data collected during pre-nursery tests revealed the quantitative nature of *Ganoderma* resistance and identified underlying loci segregating in genetic diversity that is directly relevant for the breeding program supporting the study. To assess the consistency of QTL effects between pre-nursery and field environments, information was collected on the disease status of the genitors planted in genealogical gardens and modeled with pre-nursery-based QTL genotypes. In the field, individuals were less likely to be infected with *Ganoderma* when they carried more favorable alleles at the pre-nursery QTL. Our results pave the way for a MAS of *Ganoderma* resistant and high yielding planting material and the provided proof-of-concept of this efficient and cost-effective approach could motivate similar studies based on diverse breeding programs.

Keywords

Oil palm, basal stem rot disease, nursery screening test, breeding population, QTL mapping, pedigree-based analysis.

Acknowledgments

This study was based on a very intensive and laborious work involving many people in the long-term. We thank Zulkifi Lubis, Augustiaman Purba, Shri Jeweyen, and all the SOCFIN Indonesia staff at Tanah Gambus who

performed the pre-nursery trials. We thank the PalmElit staff, Hubert de Franqueville and Michaël Pernaci for information on plant pathology and Nicolas Turnbull on breeding. We acknowledge Tristan Durand-Gasselin (PalmElit) for his insightful review of the study and manuscript. We thank Eric van de Weg (Wageningen UR) for the review of the manuscript.

This research was partly funded by a grant from PalmElit SAS. MD contributed partly to this study while she was visiting researcher at Georgetown University and supported by the European Union's Horizon 2020 research and innovation program under grant agreement No840383.

CRediT authorship contribution statement

Conceptualization: Benoît Cochard, Sébastien Tisné ; Data Curation: Sandrine le Squin, Virginie Riou, Florence Jacob ; Formal analysis: Aurélie Daval, Sandrine le Squin, Sébastien Tisné; Investigation: Virgine Pomiès, Frederic Breton, Nopariansyah; Methodology: Sandrine le Squin, Marie Denis, Sébastien Tisné; Project administration: Florence Jacob, Sébastien Tisné ; Resources: Nopariansyah, Benoît Cochard ; Software: Marco Bink ; Supervision: Florence Jacob, Norbert Billotte, Sébastien Tisné ; Writing - Original Draft: Aurélie Daval, Sébastien Tisné ; Writing - Review & Editing: Aurélie Daval, Marie Denis, Frederic Breton, Benoît Cochard, Florence Jacob, Norbert Billotte, Sébastien Tisné ; Visualization: Aurélie Daval, Sébastien Tisné.

1. Introduction

The African oil palm (*Elaeis Guineensis* Jacq.) is the leading oil crop worldwide with a global annual production of around 73 Mt, and accounts for more than 35% of all the edible vegetable oil produced worldwide (USDA statistics, 2019). Oil palm is expected to be able to respond to the global increase in the demand for vegetable oil projected to be 240 Mt in 2050, even higher if its non-food uses are included (Corley 2009). The oil palm sector has agreed on sustainability goals to reach this global demand (Rochmyaningsih 2019), in particular through the certification of sustainable produced palm oil (the Roundtable on Sustainable Palm Oil, RSPO, https://rspo.org/). However, pests and diseases threaten palm oil production in all areas of cultivation and contribute to the current yield gap (Woittiez et al. 2017). If it is to achieve the zero-deforestation goal in high conservation value forests included in the RSPO commitments, oil palm will inevitably be cultivated on existing arable lands under increasing pathogen pressure. The integrated pest management (IPM) covers sustainable solutions to this problem including improved plant disease resistance. Oil palm breeders thus needs to focus on developing resistant planting material, while maintaining or even improving oil yield.

The basal stem rot disease caused by *Ganoderma boninense* is a major threat in South East Asia, with projections worsening due to climate change (Paterson 2019). This pathogenic fungus is a soil-borne basidiomycete that mainly infects the oil palm when its roots come into contact with infected debris or with the roots of neighbor palms (Rees et al. 2009). *Ganoderma* stem rot disease has a significant effect on oil yield even when only 10-20% of palm trees are infected, and 30-70% of the trees may have died over a typical 25-year planting cycle (Durand-Gasselin et al. 2005; Cooper et al. 2011). To date, no specific interaction and/or complete resistance have been identified in oil palm/*Ganoderma* pathosystem, which is consistent with its hemibiotrophic pathogenic lifestyle. However, observations of contrasted levels of resistance in diverse genetic backgrounds suggest that breeding for quantitative disease resistance (QDR) is a promising solution (Franqueville et al. 2001). Typically, research on perennial plant disease resistance is based on large scale costly field experiments, even more so when investigating QDR. When possible, *ex situ* experiments with controlled inoculation of the pathogen are powerful tools that offer more repeatability and increase both speed and throughput, especially in genetic surveys. In oil palm, such pre-nursery tests were first developed for research on vascular wilt (De Franqueville and Renard 1990), followed by *Ganoderma* in the 2000s (Idris et al. 2004; Breton et al. 2006b; Rees et al. 2007) and are now widely used. However, transferring results to the field can be problematic because of a more complex biotic context, the age specificity of the QDR mechanisms, or the effects of cultural practice management on disease epidemiology. Despite these challenges, by combining field and pre-nursery approaches in long-term works in the framework of an oil palm breeding program, Cirad, its subsidiary PalmElit, and their partners have managed to release planting material that is highly resistant to vascular wilt and intermediate resistant to basal stem rot caused by *Ganoderma* (De Franqueville and Renard 1990; Franqueville et al. 2001; Durand-Gasselin et al. 2005; Breton et al. 2009).

Information on the genetic architecture and molecular determinisms of traits of interest could help shorten the long breeding cycle of oil palm, which currently exceeds 20 years, and would be particularly useful in the case of *Ganoderma* disease given the cumbersome nature of field and nursery trials. Marker assisted selection (MAS) based on this information would increase the annual genetic gain thanks to both accelerated evaluation of selection candidates and increased selection intensity by enabling surveys of wider genetic diversity at the same cost (Cros et al. 2015, 2017). Moreover, identification of the genetic bases of resistance to *Ganoderma* could resolve the challenge of breeding for both QDR and yield related traits (Nelson et al. 2018) by using simulation and prediction tools (Tisné et al. 2019). Most molecular studies on *Ganoderma* disease to date have been based on inoculated *vs* non-inoculated seedlings at the pre-nursery stage, with no or low genetic diversity. The first investigations focused on *a priori* selection of candidate resistance genes to fungal diseases (Yeoh et al. 2012, 2013; Tan et al. 2013).

Next the genes, proteins and pathways affected by *Ganoderma* infection were identified using broader transcriptomic (Tee et al. 2013; Ho et al. 2016; Bahari et al. 2018; Faizah et al. 2020; Sakeh et al. 2020), proteomic (Al-Obaidi et al. 2014) and metabolomic (Nusaibah et al. 2016) approaches. Considering that *Ganoderma* is a white rot fungus (Paterson 2007), lignin related traits were investigated as putative QDR mechanisms by surveying the response of lignin content and composition to *Ganoderma* infection together with the associated genes (Govender et al. 2017). Lignin related traits and nutritional traits were found to differ in progenies with different levels of resistance to *Ganoderma* (Govender et al. 2020) but the restrained genetic design confounds the effects of genetic and resistance variation.

QTL mapping offers an alternative approach that provides information on the genetic architecture based on a relevant genetic diversity, with no *a priori* biological knowledge. The detected loci form the basis of the MAS strategy but also provide insights into the mechanisms and genes involved in the QDR. The first published QTL study reported the analysis of 79 individuals from one resistant and two susceptible families based on 58 simple sequence repeat markers and found alleles associated with *Ganoderma* symptoms (Hama-Ali et al. 2015). More conclusive insights would require much more data, but QTL analyses of oil palm crosses are typically not sufficiently effective due to biological and cost constraints (Jeennor and Volkaert 2014; Lee et al. 2015; Pootakham et al. 2015). This is even more problematic for field studies that are indispensable to assess genetic diversity in an agronomic context, whose implementation is very costly and would result in lower production income due to the disease context. A powerful and cost-effective approach is to directly use the databases compiled in ongoing breeding programs, which are typically large and obtained from diverse relevant genetic backgrounds, to map *in silico* the QTLs for the traits of interest (Parisseaux and Bernardo 2004). Despite the potential of this approach, data from breeding programs are unique, mainly because of a complex genetic design that may be biased due to selection, or unbalanced phenotyping coverage. Thus, they require appropriate statistical models for their development and evaluation in contrasted contexts, which are currently an active research topic (Würschum 2012; Garin et al. 2017; Korontzis et al. 2020). In oil palm, an *in silico* QTL mapping approach based on the two step variance component approach considering identity by descent (IBD) information (George et al. 2000; van Eeuwijk et al. 2010) yielded promising results on production traits recorded in large scale evaluation genetic trials (Tisné et al. 2015). This approach was successfully extended to survival data and applied to a multi-parent population to detect *Ganoderma* resistance QTLs in the field, allowing to identify two QTL related to the occurrence of the first disease symptoms, and two related to the death due to *Ganoderma* (Tisné et al. 2017). A Bayesian approach to pedigree based QTL mapping using IBD information was also developed in the 2000s and implemented in the

FlexQTL software (van de Weg et al. 2004; Bink et al. 2008). This made it possible to carry out increasing numbers of studies in several crops that share the constraints and potential described above for oil palm, in particular for disease resistance in strawberry (Mangandi et al. 2017; Anciro et al. 2018) or in apple (van de Weg et al. 2018).

In this study, we evaluated the potential of an *in silico* approach based on the large existing databases of a long-term oil palm breeding program for the study of *Ganoderma* resistance. We genotyped an existing DNA bank primarily established for identity checking purpose and performed a pedigree-based QTL mapping using data recorded in *Ganoderma* pre-nursery trials over a period of more than ten years. We then assessed the consistency of pre-nursery QTL effects in natural field conditions using a database recording the *Ganoderma* infection status over years for the palms planted in genealogical gardens. Thus, using a cost-effective approach that is directly relevant to the breeding program, we were able to study two major issues, i.e. the genetic architecture and consistency between pre-nursery and field results, paving the way for the implementation of MAS for *Ganoderma* resistant planting material.

2. Material and methods

2.1. Plant material

The palm trees used in this study belong to the oil palm breeding program of Cirad, its subsidiary PalmElit and their partner PT Socfin Indonesia (Indonesia). This breeding program is conducted in a recurrent reciprocal selection scheme with two heterotic groups A and B (GA and GB to produce superior GA×GB hybrid crosses used as commercial planting material (Gascon and De Berchoux 1964; Meunier and Gascon 1972). Individuals from different heterotic groups have complementary yield component traits, with low fruit bunch number and high bunch weight in GA and reciprocally in GB. GA×GB hybrids consequently show a heterosis effect on fruit bunch yield. Moreover, individuals included in GA are Dura palms, homozygous for the thick alleles of the shell gene (Singh et al. 2013) while individuals included in GB are Pisifera (homozygous alternative alleles), the hybrid GA×GB being Tenera which is the most productive form with thin shell. The parental population studied for the *Ganoderma* resistance included only individuals from GB, grouping genetic origins of La Mé (LM, Ivory Coast) and Yangambi (YBI, Republic Democratic of Congo). The GB pedigree used in the pre-nursery analysis comprised 372 individuals including founders, with 246/126 from LM/YBI genetic origin respectively and 240 /93 genotyped (Supp. Table 1). Among them 200 LM and 83 YBI parents were directly progeny tested for *Ganoderma* resistance in a pre-nursery screening test (Fig. 1). The individuals were distributed over many full-sib families derived from a small number of founders through consecutive crosses or self-pollinations in the framework of the ongoing

breeding program (Fig. 1). Among the 372 individuals in the whole pedigree, 219 LM individuals were planted between 1970s and 2000s at the same location (Bangun Bandar, Indonesia) and were used for subsequent field analysis.

2.2. Phenotypic data

2.2.1. Pre-nursery screening tests

An early pre-nursery screening test was developed in the 2000s by Cirad and Socfin Indonesia in the Tanah Gambus estate, Indonesia. The first objective was to speed up the evaluation of genetic resistance to *Ganoderma* of commercial oil palm planting materiel, using controlled and standardized inoculation of germinated seeds (Breton et al., 2006). The inoculation of germinated seed was performed using a 12 week-old *Ganoderma*-colonized rubber wood block (108 cm3) as inoculum source, that was previously deposited in the nursery polybag before the seeds were planted.

A pure dikariotic *Ganoderma boninense* isolate was used in all the trials (NJ), previously harvested from an infected oil palm planted in Bangun Bandar, SOCFINDO estate (Mercière et al. 2015). This isolate was successively regenerated from the bole of young infected seedlings in consecutive pre-nursery trials to provide several dikariotic clonal lines (CL, n=7) over the 10 years of testing. These reactivating steps of the isolate made it possible to avoid the loss of pathogenicity often observed after successive sub-cultures on artificial fungi growth media (Butt et al. 2006). A single pathogen CL was used for all the crosses tested in a single trial. Around 100 crosses were assessed simultaneously in each pre-nursery trial. Among them, 20% were control crosses from susceptible, intermediate and resistant genetic backgrounds and were included in all the trials performed. Of the remaining 80% of crosses representing the tested crosses, 50% overlapped two consecutive trials, leading to at least two independent tests per tested cross. Each cross was represented by 100 inoculated germinated seeds clustered in five replicates following the protocol described by Breton et al. (2009). Inoculated seedlings were observed every four weeks for the appearance of the first external disease symptom, on average between 8 and 12 weeks after inoculation of the germinated seeds, after which the disease symptoms were recorded at two weekly intervals as (1) infected and (0) if not infected. The trial was stopped when the average percentage of infected seedlings within the group of control crosses reached 30%, usually around 34 weeks after inoculation of the germinated seeds. This 30% threshold was determined to have the best “discriminating power” between the resistant and sensitive control crosses, and so among the tested progenies (Breton et al. 2009).

This study included 102 *Ganoderma* pre-nursery screening test trial, covering 10 years of data recording. The trials performed between 2007 and 2017 represented the evaluation of 4,017 unique crosses, from either GA×GA, GA×GB or GB×GB genetic background. Considering that the purpose of this study was to assess the genetic bases of *Ganoderma* resistance in the commercial genetic material, only the GA×GB crosses were taken into consideration (n=3,792), derived from 2,037 and 340 individuals from the GA and GB respectively. Each parent from GB included in the analysis was progeny tested in an average of 20.5 GA×GB crosses.

2.2.2. Statistical modeling of pre-nursery data

The resistance of the GB individuals was progeny-tested through several GA×GB crosses involving them as GB parents. The response variable Y considered in this study was the proportion of affected progenies per cross at the end of the trial. A first step of statistical modeling of Y was necessary to obtain a single value per genotype required for the QTL analysis while accounting for nuisance effects due to the long-term data. Y was modeled using generalized linear mixed models (GLMM). Briefly, in a GLMM, Y is assumed to be generated by a particular distribution in the exponential family. The conditional mean of the distribution μ is linked to a linear predictor η which contains fixed and random effects, through the inverse link function g^{-1}:

$$g(\mu) = \eta = X\beta + Z_T u_T + Z_A u_A + Z_B u_B + Z_C u_C$$

where X is a $n \times m$ design matrix relating observations to *Ganoderma boninense* CL fixed effects β where β is a $m \times 1$ vector ($m = 7$), Z_T is a $n \times t$ design matrix relating observations to trial random effects $u \sim N(0, I\sigma_T^2)$ with u is a $t \times 1$ vector ($t = 102$), Z_C is a $n \times c$ design matrix relating observations to specific combining ability (SCA) random effects $g_C \sim N(0, I\sigma_c^2)$ where g_C is a $c \times 1$ vector ($c = 3{,}792$), Z_A and Z_B are $n \times q_A$ and $n \times q_B$ design matrices relating observations to general combining ability (GCA) random effects for GA and GB, $g_A \sim N(0, A_A\sigma_A^2)$ and $g_B \sim N(0, A_B\sigma_B^2)$ respectively, where g_A and g_B are $q_A \times 1$ and $q_B \times 1$ vectors, respectively ($q_A = 2{,}037$ and $q_B = 340$). A_A and A_B are the pedigree-based kinship matrices of GA and GB, respectively.

In our work, we explored two types of distributions: binomial distribution, which is the appropriate one for proportional data, and normal distribution, for which more derived genetic parameters can be estimated.

The first model considers a binomial distribution such as:

$$Y_{c,t} \mid u_t, u_A, u_B, u_C \sim Bin\,(n_{c,t}, \pi_{c,t})$$

where $Y_{c,t}$ is the number of affected progenies in the cross (c) and the trial (t) among the number of inoculated progenies $n_{c,t}$, and $\pi_{c,t}$ is the associated probability.

The link function g is the logit such as:

$$g(\pi_{c,t}) = \log\left(\frac{\pi_{c,t}}{1-\pi_{c,t}}\right) = \eta_{c,t}$$

The second model considers a normal distribution such as:

$$Y_{c,t} \mid u_T, u_A, u_B, u_C \sim N\,(\eta_{c,t}, \sigma^2)$$

where $Y_{c,t}$ is the proportion of affected progenies in the cross (c) and the trial (t) , σ^2 is the residual variance, and the link function is the identity. Note that this second model is a linear mixed model (LMM).

Both models enabled prediction of the best linear unbiased predictor (BLUP) for each GB individual used in the QTL mapping, A_B being replaced by an identity matrix in order to avoid using the pedigree information that was subsequently used in the QTL analysis. Both statistical models were performed using ASReml-R software (Butler et al. 2007, V4) and resulted in two vectors of BLUP for group B individuals that were used in subsequent QTL mapping analysis.

2.3. Molecular data and genetic map construction

The 334 freeze-dried oil palm leaf samples available at the Cirad DNA-bank for the GB individuals included in the analysis were genotyped with 199 SSR markers developed in different studies. Among the 199 markers, 177 markers were developed by Cirad (Billotte et al. 2005), two by the Lee et al. (2015), four markers by the Malaysian Palm Oil Board (MPOB) (Zaki et al. 2012) and 18 expressed sequence tags markers were developed by IRD (*Institut de Recherche pour le Développement*) and Cirad (Tranbarger et al. 2012). These markers were selected based on a previous integrated pedigree-based genetic map constructed from a population of related individuals (Cochard et al. 2015). Selection was for a uniform distribution in the genome and the highest level of polymorphism in both LM and YBI genetic backgrounds. The information concerning markers was gathered in the supp. Table 2. DNA extraction, evaluation of the DNA concentrations and microsatellite fragment amplification were performed using the protocol described in Cochard et al. (2015). Genemapper© V4.1 (Applied Biosystems, USA) software was used to determine the size of the alleles.

Three genetic maps were constructed, one for each of LM and YBI population and one integrated map using the pedigree-based linkage mapping software CRI-MAP v2.4 (Green et al. 1990), as described in Cochard et al. (2015).

Consistency of marker calling across pedigrees and absence of spurious rates of double recombination events were checked using both CRI-MAP and FlexQTL™, and data were improved where necessary. Genetic maps were drawn using MapChart v2.0 software (Voorrips 2002) and are presented in Supporting Information Figure S1.

2.4. Pre-nursery QTL mapping approach

QTL mapping of *Ganoderma* disease resistance in pre-nursery conditions followed two main steps. The first step was carried out using a Bayesian approach and a multiple QTL model implemented in FlexQTL™ (Bink et al. 2002, 2014, 2008; www.flexqtl.nl) on the pre-nursery data after modeling, in order to identify putative QTL positions and predict the QTL genotypes. The second step consisted in stepwise QTL model selection on the raw pre-nursery data using the predicted QTL genotypes as fixed effects in the LMM.

2.4.1. QTL region identification and QTL genotype prediction

Six separate QTL analyses, corresponding to the two vectors of GB individual BLUP (see Phenotypic data section) with three different starting random seeds were performed using FlexQTL™. The six QTL analyses were based on a model with additive QTL effects, with the parameters MaximQTL and priorQTL set at 20 and 5 respectively for the Markov chain Monte Carlo simulation. The length of the Markov chains were set at 1 000 000 with a thinning value of 1 000. Using these parameters, the convergence indicators reached satisfying values for each parameter assessed (overall mean, μ, the residual variance, σ_e^2, the number of QTLs, $N_{QTL,}$ and the variance of QTLs, v_{QTL}). QTL regions were marked from the marginal posterior distributions of the six simulations and consensus QTL positions identified at the peaks of the summed posterior intensities profiles over the six simulations. QTL regions were named by the concatenation of population ID (LM, YBI or GB which refers to the grouped LM and YBI populations), the linkage group and the peaks separated by "@". For each consensus QTL, QTL genotypes for all individuals in the pedigree were predicted based on the vectors of QTL genotype posterior probabilities extracted from the FlexQTL output "MQTRegionsGTP.csv". QTL genotypes values were calculated as $\left[\left(0 * P_{qq}\right) + \left(1 * P_{Qq}\right) + \left(2 * P_{QQ}\right)\right]$, with P the probability associated with the qq, qQ and QQ QTL genotypes, q being the favorable allele in this case. The continuous [0,2] values of the QTL genotypes were converted into discrete values {0,1,2} using the following threshold: values in the ranges [0,0.7],]0.7,1.3[and [1.3,2], were assigned to 0, 1 and 2 respectively, corresponding to individuals carrying homozygous favorable, heterozygous or homozygous unfavorable disease resistance alleles at the respective QTL regions considered.

2.4.2. Stepwise QTL model selection

In order to obtain a full QTL model fitted on the raw phenotypic data, QTL results from different modeling and random seeds were aggregated using stepwise model selection. The stepwise approach was applied on QTL genotypes vectors tested in the LMM model (see Phenotypic data section), following the procedure of the stepwiseqtl function of the R/qtl package (Broman and Sen 2009). First, a main effect QTL model was selected by testing the QTL genotype vectors in the LMM model with sequentially, a forward selection and a backward elimination. Model selection was based on the Akaike information criterion (AIC, Akaike 1998) using the full loglikelihood (Verbyla 2019). Similarly, the main effect QTL model was extended to the complete QTL model by first testing the interactions between QTLs and both QTL and CL (fixed effects), and second with the GA genetic background (random effect). Stepwise model selection was performed using ASReml-R software (Butler et al. 2007, V4).

2.5. Field evaluation of pre-nursery QTL

The relationships between *Ganoderma* genetic resistance in pre-nursery and field conditions were investigated using the census of disease status of the La Mé parents planted in genealogical gardens (see plant material section). The *Ganoderma* infection status was recorded biannually on 219 LM individuals planted in 1974 (5), 1976 (11), 1996 (5), 1997(107), 1998 (1), 1999 (47), 2001 (20) and 2003 (23) in six different blocks at Bangun Bandar estate, Indonesia. The disease status recording began within the three years after planting in the case of plantation after 1990 and in the 2000s for older plantings, and the last observation was recorded in 2018. *G. boninense* disease symptoms were scored blindly based on a six-level scale as described in Tisné et al. (2017). The appearance of the first *Ganoderma* symptom (T1S, first observation of score 2–6) was recorded and the associated time was considered as survival time, i.e., time from planting to the time the event occurred. The survival data were analyzed using the Cox model integrating a fixed effect for the date of planting:

$$\lambda(t, X) = \lambda_0(t) e^{X\beta} \quad (1)$$

where t is the time to the event or censoring, λ_0 denotes the baseline hazard function, X is the $n \times d$ design matrix relating the survival outcome for individuals to date of planting effects ($d = 8$) and $\beta = (\beta_1, \dots, \beta_d)$ is a $d \times 1$ unknown vector.

The effects of pre-nursery QTL were evaluated using the likelihood ratio test, for which the limiting distribution follows a chi-squared distribution, between the model (1) and the following model (2):

$$\lambda(t, X) = \lambda_0(t) e^{X\beta + X_q q} \quad (2)$$

with X_q being the {0,1,2} vector of pre-nursery-based QTL genotypes for the individuals and q the QTL effect. The analysis was performed with R software version 3.2.3 (Team 2012) and the *survival* package (Therneau 2015).

3. Results

3.1. Segregation of *Ganoderma* resistance in the GB population

Resistance to *Ganoderma* disease was tested in pre-nursery trials on 3,792 GA×GB crosses. On average, 30.8% of oil palm seedlings per cross presented disease symptoms at the end of the trial, ranging from 3 to 92.5% among the different crosses (Fig. 2a). Both LMM or GLMM models led to very similar predictions of GCA for the GB parents (r=0.97). Predictions of GCA were higher in YBI genetic background compared to LM, indicating higher susceptibility of the YBI background tested in this study (Fig 2b-c). Within genetic backgrounds, the distribution of GCA indicated segregation of quantitative resistance among founders, with mainly additive effects. Indeed, in LM genetic background, LM_1 self-pollinated individuals were the most resistant, and all the combinations of LM_1 and the alternative founders LM_2 or LM_3 showed higher resistance than the populations derived from self-pollinations of LM_2 and LM_3 (Fig. 2b-c). Similarly in YBI, YBI_3 was the least resistant genetic background, but its combination with YBI_2 improved the resistance of derived individuals. Even in narrow genetic bases, i.e. self-pollinated progenies of the most recent generation, there was still segregation of the resistance supporting the quantitative nature of *Ganoderma* resistance (Fig. 2b-c).

3.2. Genetic bases of *Ganoderma* resistance in pre-nursery trials

QTL mapping of the *Ganoderma* disease resistance in the GB population was performed using a Bayesian approach. Cumulating both modeling and the three random seeds per model, the number of QTLs was 125 considering all the marked QTL regions found by FlexQTL, regardless the 2lnBF threshold (supp. Table 3). These 125 QTL corresponded to around 20 QTLs on average per simulation. The QTLs were distributed in 30 consensus regions covering every linkage group (LG), with overall, a similar pattern between the different simulations (Fig. 3). Among these 29 QTL regions, 11 located on LG 1, 5, 6, 8, 9, 10, 12, 13 and 16 were identified consistently in the six simulations. The QTL mapping performed separately in LM and YBI revealed different QTL patterns between them: consistent QTL regions on LG 1, 6, 10, 12 and 13 segregated in the LM genetic background while the regions were located on LG 5, 8, 9 and 10 in the YBI genetic background (Supporting Information Figure S2). The average length of the QTL interval was around 25 cM (4-107 cM). Considering QTL genotypes in the 30

consensus QTL regions, there were on average, 35, 41 and 24% of *QQ*, *Qq* and *qq* genotypes respectively, in the GB population, q being the favorable allele in this case.

Stepwise model selection was performed based on the QTL genotype vectors calculated for the 30 consensus QTL regions. The first step fitted the LMM and indicated that the components related to the genetic effects represented 21% of total phenotypic variation, while 6% corresponded to the GCA of the GB individuals (Fig. 4). The final QTL model retained four main effect QTL on LG 8, 9, 10 and 16, and one in interaction with the GA genetic background on LG 6 (Fig. 4, supp. Table 4). Adding either the main effect or interacting QTLs in the LMM in the different steps did not change the values of the non-genetic components, whereas the GCA_{GB} was reduced to 1%. Including the interaction between the QTL on LG6 and the GA genetic background reduced both the values of the SCA and the GCA_{GA} components. The partial determination coefficients computed for each QTL ranged from 0.05-2% of the total phenotypic variance, corresponding to 3-9% of genetic variance.

3.3. Effects of pre-nursery-based QTL on field *Ganoderma* resistance in the La Mé parents

The effects of the QTL identified using the pre-nursery data on GA×GB crosses were evaluated in the field where 219 LM parents included in the pre-nursery study were planted and underwent natural, uncontrolled *Ganoderma* infection. The time of the first *Ganoderma* symptom appearance (T1S) was modeled using Cox regression with the date of planting as covariate ($P < 0.01$). The effect of the percentage of favorable alleles per individual among the 21 QTL regions identified in the LM genetic background (range 28-75%) was first assessed to evaluate the global trend between pre-nursery and field conditions. The percentage of favorable alleles effect was not found to be significant (P=0.2), but Kaplan-Meier estimates of survival showed consistency between the pre-nursery and field QTL effects, a higher percentage of favorable alleles increased the probability of survival (Fig. 5a). Hence, the individuals with less than 50% of favorable alleles were twice more affected by *Ganoderma* 20 years after planting than individuals with more than 50% of favorable alleles (Fig. 5a). Then QTL genotype vectors, predicted either GB or LM populations, were tested one at a time as covariates in the Cox model. The level of statistical evidence of QTL effects between pre-nursery and field data was not correlated and significant QTL effects were found for both a high (LG 9) or low (LG 4, 15) level of evidence in pre-nursery conditions (Fig. 5b). However the direction of effects between field and pre-nursery effects was consistent for 78% of the QTLs, and for 89% when a P-value=0.05 threshold was applied in the Cox model (Fig. 5b, Supporting Information Figure S3).

4. Discussion

Marker assisted selection (MAS) has a great potential for plant breeding and has been widely used for many crops with substantial achievements, especially for resistance to biotic stresses (Muranty et al. 2014). MAS should be particularly useful for perennial crops with a long breeding cycle and high phenotyping costs like oil palm, despite the identified biological, socioeconomic or technical issues (Muranty et al. 2014). In this paper, we report the proof of concept of an efficient *in silico* QTL mapping approach based on data collected in an ongoing breeding program. This allowed us to gain valuable insights into the genetic architecture of *Ganoderma* resistance and the transferability between field and pre-nursery results, as a basis for a future MAS.

4.1. Opportunities and issues of QTL mapping using data from breeding programs

Breeding programs for perennials are inherently geared towards long-term work with extensive data recording. This make them highly suited to the *in silico* approach, which is likely to improve the statistical properties of QTL detection through the increase in population size and diversity compared to conventional biparental populations. However, the specificity of the data from breeding programs, such as the extent of non-genetic effects due to long-term data or the genetic and phenotypic design unbalances due to the selection process, could reduce the expected benefits of QTL detection, namely its power and the accuracy of QTL location and QTL effect estimation (Würschum 2012). Hence, these datasets require a first stage of statistical modeling to account for several non-genetic effects and to obtain genotypic values. Thanks to their flexibility, mixed models are ideal tools to handle several types of data and effects (Smith et al. 2005). We used two types of mixed models, LMM and GLMM that enabled us to predict the GCA of genotyped individuals while accounting for confounding effects. We subsequently used these GCA values in FlexQTL because this software requires only one value per genotyped individual whereas they were progeny tested in the pre-nursery trials. Such a two-stage approach could affect QTL results so one-stage approaches are preferred when possible (Xue et al. 2017; Barrasso et al. 2019). The two types of mixed model used in this study did not lead to major differences in the QTLs identified, and a one-stage IBD–based variance component approach previously reported for production traits (IBD-VC, Tisné et al. 2015) that we used on pre-nursery *Ganoderma* data also produced similar results (data not shown). However, the calculation time requirement for the IBD-VC is an obstacle to a proper estimation of the significance threshold by permutation and a multi-QTL mapping procedure, which made us favor the approach presented.

Few studies have assessed the effects of the dataset features on QTL detection. In barley, using GWAS with an unbalanced dataset, the false positive rate was increased, whereas one-stage analysis performed better (Wang et al. 2012). In durum wheat, a GWAS performed both on an unbalanced and balanced dataset from a breeding

program showed major overlapping of selected SNP (Johnson et al. 2019). In diploid potato, a dataset grouping F3 families under selection was analyzed using either GWAS, stratified linkage or IBD based approaches that led to consistent QTL detection, but revealed issues concerning the QTL allele frequencies that could affect the results (Korontzis et al. 2020). In our study, the population studied could be genetically biased due to prior selection of the crosses tested for *Ganoderma* resistance based on yield related traits. However, inspection of QTL genotype frequencies showed that there were no depleted allelic classes among the QTL retained in the stepwise model selection. Moreover, the QTL genotype vectors predicted at the QTL regions were not correlated for the different linkage groups, indicating little segregation distortion that could have arisen due to the selection process.

Concerning the accuracy of QTL location, the increased population size allowed by the *in silico* approach should reduce the QTL interval thanks to the increased number of recombinations. In this proof of concept study, we chose to genotype the population using well characterized SSR markers in order to be able to connect the results with previous ones obtained with related populations. However, the QTL intervals were much larger than in other studies using FlexQTL on populations of similar size but with thousands of markers, indicating that the density was insufficient to mark them accurately. The large QTL regions could probably be considerably reduced thanks to the favorable genetic design and we are currently performing high-density SNP genotyping to achieve this objective. Beyond this limitation, the use of FlexQTL was particularly interesting: the use of IBD information mitigates the effect of low density genotyping, and the prediction of QTL genotypes offers the opportunity to use them in subsequent analyses. Hence, we were able to select a full QTL model using the raw data by testing main and interaction effects, and to assess the effects of pre-nursery QTL in the field. As reported by Verma and Whitaker (2018), QTL genotypes have great potential in the breeding context, for example, to predict QTL alleles for unobserved individuals in the breeding program based only on their marker and pedigree information, and then their expected resistance level.

4.2. Insights into the genetic architecture of *Ganoderma* resistance in oil palm

A first insight into genetic architecture came from the variance decomposition using the sire and dam mixed model designed for the analysis of the data on GA×GB hybrids. The genetic component, i.e. GCA in both heterotic groups and SCA, represented around 20% of the total phenotypic variance, which was expected due to the consistent genetic resistances identified in contrasted crosses or clones, balanced by the moderate repeatability of the screening tests (Durand-Gasselin et al. 2018). More surprising, the variance assigned to the GA pedigree was double that for the GB pedigree, while the pure parental GB genetic backgrounds are both more resistant and

exhibit more resistance variability than GA backgrounds (Durand-Gasselin et al. 2018). This could be an artefact of the unbalanced number of parents screened between heterotic groups and further investigation is needed to accurately estimate their relative contribution to the GA×GB resistance. The variance associated with SCA effect was 20% of the genetic variance and one QTL×genetic background interaction was retained, while well supported previous observations indicated that resistance was mainly additive, both in pre-nursery and field trials (Durand-Gasselin et al. 2018). Again, this could be an artefact, as only the GB pedigree was genotyped for this study but further analyses using both heterotic groups will allow us to estimate the proportion of variance due to GA×GB interaction and identifying underlying QTL.

The distributions of the GCA of GB individuals showed segregation of the *Ganoderma* resistance throughout the pedigree, even in the most inbred generations. Consequently, we identified a large number of putative QTL regions using FlexQTL, with weak to moderate effects. This partially reflects the composition of the GB that grouped two contrasted populations, LM and YBI, which displayed distinct QTL patterns when analyzed separately. However, even when we focused on a restricted genetic background, the large number of putative QTL found despite the reduced population size confirm the quantitative nature of *Ganoderma* resistance (quantitative disease resistance, QDR). Thus, the marked difference in *Ganoderma* resistance consistently observed between the four full-sib founders of the studied LM pedigree (Durand-Gasselin et al. 2018) is rather the consequence of a better combination of many favorable alleles than of a limited number of major QTLs. The numerous QTL found and the dissimilarity of QTL patterns between the LM and YBI genetic backgrounds is likely due to either the *Ganoderma* bio-trophic pathogenesis that induce contrasted transcriptomic responses (Bahari et al. 2018) or the multiple mechanisms involved in the QDR (Poland et al. 2009). This could explain the few discrepancies observed for some pre-nursery QTL with no effect in the field, and even a QTL with an opposite effect on LG12, considering that such QDR mechanisms are more prone to depend on the age of palms, on the environmental conditions, or on the genetic background surveyed.

Inspection of QTL colocalization may validate putative QTL when found for similar traits in independent experiments and inform QTL pleiotropy or linkage for different traits. Pleiotropy is especially worth investigating for QDR to obtain insights into possible underlying mechanisms and, together with linkage, on the resulting trade-off with other traits of interest (Nelson et al. 2018). To date, only two genetic mapping studies have been reported on *Ganoderma* resistance. The first analyzed data from a nursery test involving one resistant and two susceptible progenies, with a similar genetic background (Deli×YBI) and common markers to our study (Hama-Ali et al. 2015). Despite the limited scope of the study, i.e. involving only 79 individuals genotyped with 58 SSRs, Hama-

Ali et al. (2015) identified two significant markers on LG2 and seven in the same QTL regions as in our study, what is more, in equivalent populations, YBI and GB respectively. The second study used field data recorded on a multi-parental GA×GB population involving four GB founders that were the same as in the present study (*Eg*9PP population, Tisné *et al.* 2017). Four *Ganoderma* resistance loci were identified, two controlling the occurrence of the first *Ganoderma* symptoms (T1S), and two the death of palm trees (TD). Among them, the T1S QTL at the bottom of LG1 collocated with a QTL identified in GB and LM populations in the present study. The *Eg*9PP population and a large-scale genetic trial involving GB parents related to the founder of the present study (NGP population, Tisné *et al.* 2015, Tisné *et al.* 2019) were evaluated in the framework of the breeding program. Hence, data for fruit bunch production, oil extraction rate, and height increment traits were stored in databases, and both populations as well as the population from the present study were genotyped with the same SSR markers from a reference genetic map (Cochard et al. 2015) allowing QTL detection. We observed that among the six *Ganoderma* QTL regions with higher statistical support found in the GB, LM or YBI populations, most collocated with a large number of QTL for other agronomic traits (Tisné, personal communication). The colocalizations were more frequent in the LM population (33) than in the YBI one (15), while they were mostly found with oil extraction rate related traits and bunch number in LM genetic background in contrast with bunch weight and height increment in the YBI one (Tisné, personal communication). These preliminary findings now require further support, in particular by using a high-density SNP genotyping that is currently in progress, but already provide interesting insights into the possible diverse mechanisms underlying the QDR, which could differ considering the genetic backgrounds. This also highlights the benefits of the *in silico* approach assessed in this study that makes it possible to gather information from the entire breeding program for a more comprehensive description of the genetic architecture of traits of interest.

4.3. Advances towards a MAS of *Ganoderma* resistance in oil palm breeding programs

No complete resistance to *Ganoderma* has been identified to date and the results of the present study corroborate previous observations to indicate its quantitative nature (Franqueville et al. 2001; Idris et al. 2004; Durand-Gasselin et al. 2005). Despite the increasing use of QDR to improve the sustainability of disease resistance (Poland et al. 2009; Roux et al. 2014) the high number of loci and mechanisms involved makes its selection challenging. This is more acute in the case of oil palm with its long breeding cycle, worsened by the slow *Ganoderma* disease progression. Pre-nursery testing accelerated the screening of genetic material and revealed a genetic component that accounted for about 20% of phenotypic variance, which is generally a favorable level for a MAS perspective (Muranty et al. 2014). A first concern is to insure the consistency of QTL effects between the pre-nursery and field

results, like in conventional selection (Durand-Gasselin et al. 2018). We attempted to assess this at the QTL level with the extensive use of the data from the breeding program, including the *Ganoderma* census routinely recorded on seed and genealogical gardens. Following the previous study assessing the *Ganoderma* resistance in field we used a survival analysis approach that provides several advantages (Tisné et al. 2017). Despite the limitations of specific to the data recorded in seed gardens, i.e. mature palms of pure genetic backgrounds in the field vs GA×GB seedlings in pre-nursey and spatio-temporal heterogeneity in the field, the accumulation of favorable pre-nursery QTL alleles improved field resistance. Interestingly, the majority of QTL effect directions were consistent regardless the statistical evidence in pre-nursery. Thus, the many QTL that would not have been detected in the field setup because of a lack of statistical power, were identified in the pre-nursery study and are valuable for a marker-assisted *Ganoderma* resistance selection.

Secondly, the quantitative nature of *Ganoderma* resistance identified could hamper the conventional QTL pyramiding approach due to the high number of loci involved, especially considering the long generation time in oil palm. In such a QDR context, the MAS approaches developed for other agronomic quantitative traits are probably more suitable, especially the genomic selection (GS) approach (Poland and Rutkoski 2016). In oil palm, GS has emerged as an efficient MAS method and is being increasingly evaluated for yield improvement (Nyouma et al. 2019). Thus GS statistical models and implementation modes already assessed in oil palm could be transferred or adapted to *Ganoderma* disease related data from the breeding program (Cros et al. 2015, 2017). However, the qualitative/quantitative nature of disease resistance is a continuum (Poland et al. 2009). Despite a large number of QTL regions identified using FlexQTL, only 5 QTL with weak to moderate effects explained almost all the GB GCA component based on pre-nursery data. GS models including information on QTL or genes have been proposed to improve prediction capacity in such situations (Bernardo 2014; Zhang et al. 2014) and should be considered for a GS of implementation in light of the emerging insights into the genetic architecture of *Ganoderma* resistance.

A final issue is that selection for *Ganoderma* resistance will need to be combined with resistance to other diseases and cannot be at the expense of other traits of interests. The cost of disease resistance through negative trade-off with performance or fitness was a long-lasting question in model plants but was less investigated in plant breeding (Brown 2002). In the former section, we described colocalization of *Ganoderma* resistance QTL with yield related ones, with a genetic background specificity of these complex patterns. Dealing with multiple traits and multiple genetic background is challenging and the QTL information provided by the *in silico* approach assessed in the present study is very valuable for comprehensive modeling of a MAS strategy. Hence, a recent study in oil palm

simulated the outcomes of alternative selection strategies on yield and its components based on their global genetic architecture, including the pleiotropy/linkage and phases between the underlying QTL (Tisné et al. 2019). Virtual individuals and crosses were simulated from the actual founders via meiosis simulations based on the QTL positions identified with FlexQTL, which thus integrated their recombination frequencies. The QTL genotypes predicted in FlexQTL enabled prediction of their multiple trait values and their incorporation in yield based on the QTL effects. This use of QTL genotypes is of prime interest as QTL genotypes can be predicted based on markers alone in any related individual, whether phenotyped or not. In the MAS perspective for *Ganoderma* resistance, this approach would help attenuate possible trade-offs with other traits of interest and optimize the combination of QDR from diverse genetic backgrounds.

5. Conclusion

The cost-effective and efficient *in silico* mapping approach assessed in this study has great potential for the implementation of MAS of traits of interest in oil palm. Its application in the context of *Ganoderma* disease resistance enabled us to use the considerable quantities of data generated in the framework of conventional phenotypic selection to obtain valuable information in the MAS perspective. First, important information on the genetic architecture of resistance to *Ganoderma* disease was obtained, confirming its quantitative nature and identifying the loci involved. In addition, together with other ongoing works, this study sheds light on the relationships between *Ganoderma* resistance and yield related traits that could produce undesirable trade-offs. Second, the consistency between genetic resistance in pre-nursery conditions and in the field was assessed at the QTL level and globally indicated satisfactory portability. However, a few loci deserve careful consideration due to underlying mechanisms that could lead to contrasted phenotypic expression between pre-nursery and field conditions. Finally, this proof-of-concept study provides guidelines for future works on *Ganoderma* disease resistance and should encourage oil palm breeders to use this approach to collectively acquire a better comprehension of its complex genetic architecture.

Declaration of Competing Interest

The authors declare that they have no conflict of interests.

Data availability

The datasets generated and analyzed during the current study are available from the corresponding author.

Acknowledgments

This study was based on a very intensive and laborious work involving many people in the long-term. We thank Zulkifi Lubis, Augustiaman Purba, Shri Jeweyen, and all the SOCFIN Indonesia staff at Tanah Gambus who performed the pre-nursery trials. We thank the PalmElit staff, Hubert de Franqueville and Michaël Pernaci for information on plant pathology and Nicolas Turnbull on breeding. We acknowledge Tristan Durand-Gasselin (PalmElit) for his insightful review of the study and manuscript. We thank Eric van de Weg (Wageningen UR) for the review of the manuscript.

This research was partly funded by a grant from PalmElit SAS. MD contributed partly to this study while she was visiting researcher at Georgetown University and supported by the European Union's Horizon 2020 research and innovation program under grant agreement No840383.

Supplementary Information

Supporting Information Figure S1: Genetic map of the prenursery GB, LM and YBI oil palm populations.

Supporting Information Figure S2: QTL mapping of the *Ganoderma* resistance in the prenursery LM and YBI oil palm populations.

Supporting Information Figure S3: Survival curves of the La Mé population in field conditions according to the genotypes of QTL identified based on the pre-nursery data.

References

Al-Obaidi JR, Mohd-Yusuf Y, Razali N, et al (2014) Identification of proteins of altered abundance in oil palm infected with Ganoderma boninense. Int J Mol Sci 15:5175–5192

Anciro A, Mangandi J, Verma S, et al (2018) FaRCg1: a quantitative trait locus conferring resistance to Colletotrichum crown rot caused by Colletotrichum gloeosporioides in octoploid strawberry. Theor Appl Genet 131:2167–2177. https://doi.org/10.1007/s00122-018-3145-z

Bahari MNA, Sakeh NM, Abdullah SNA, et al (2018) Transciptome profiling at early infection of Elaeis guineensis by Ganoderma boninense provides novel insights on fungal transition from biotrophic to necrotrophic phase. BMC Plant Biol 18:1-25. https://doi.org/10.1186/s12870-018-1594-9

Barrasso C, Memah M-M, Génard M, Quilot-Turion B (2019) Model-based QTL detection is sensitive to slight modifications in model formulation. PLOS ONE 14:e0222764. https://doi.org/10.1371/journal.pone.0222764

579 Bernardo R (2014) Genomewide selection when major genes are known. Crop Sci 54:68–75

580 Billotte N, Marseillac N, Risterucci A-M, et al (2005) Microsatellite-based high density
581 linkage map in oil palm (Elaeis guineensis Jacq.). TAG Theor Appl Genet Theor
582 Angew Genet 110:754–765. https://doi.org/10.1007/s00122-004-1901-8

583 Bink M, Uimari P, Sillanpää J, et al (2002) Multiple QTL mapping in related plant
584 populations via a pedigree-analysis approach. TAG Theor Appl Genet Theor Angew
585 Genet 104:751–762. https://doi.org/10.1007/s00122-001-0796-x

586 Bink MCAM, Anderson AD, van de Weg WE, Thompson EA (2008) Comparison of marker-
587 based pairwise relatedness estimators on a pedigreed plant population. Theor Appl
588 Genet 117:843–855. https://doi.org/10.1007/s00122-008-0824-1

589 Bink MCAM, Jansen J, Madduri M, et al (2014) Bayesian QTL analyses using pedigreed
590 families of an outcrossing species, with application to fruit firmness in apple. Theor
591 Appl Genet 127:1073–1090. https://doi.org/10.1007/s00122-014-2281-3

592 Breton F, Hasan Y, Hariadi, et al (2006a) Characterization of parameters for the development
593 of an early screening test for basal stem rot tolerance in oil palm progenies. In:
594 Technol. Breakthr. Commer. Way Forw. Proc. PIPOC 2005 Int. Palm Oil Congr.
595 Agric. Biotechnol. Sustain. 25-29 Sept. 2005 Petaling Jaya Malays.
596 http://agritrop.cirad.fr/543369/. Accessed 31 Jan 2018

597 Breton F, Hasan Y, Hariadi S, et al (2006b) Characterization of parameters for the
598 development of an early screening test for basal stem rot tolerance in oil palm
599 progenies. J Oil Palm Res 2006:24–36

600 Breton F, Rahmaningsih MR, Lubis Z, et al (2009) Early Screening Test: A Routine Work to
601 Evaluate Resistance/Susceptibility Level of Oil Palm Progenies to Basal Stem Rot
602 Disease. In. MPOB International Palm Oil Congress (PIPOC 2009), Kuala Lumpur, 9-
603 12 November 2009. MPOB.

604 Butt TM, Wang C, Shah FA, Hall R (2006) Degeneration of entomogenous fungi. In: An
605 ecological and societal approach to biological control. Springer, pp 213–226

606

607 Cochard B, Carrasco-Lacombe C, Pomiès V, et al (2015) Pedigree-based linkage map in two
608 genetic groups of oil palm. Tree Genet Genomes 11:1–12.
609 https://doi.org/10.1007/s11295-015-0893-7

610 Corley RHV (2009) How much palm oil do we need? Environ Sci Policy 12:134–139.
611 https://doi.org/10.1016/j.envsci.2008.10.011

612 Cros D, Bocs S, Riou V, et al (2017) Genomic preselection with genotyping-by-sequencing
613 increases performance of commercial oil palm hybrid crosses. BMC Genomics
614 18(1):1-17. https://doi.org/10.1186/s12864-017-4179-3

615 Cros D, Denis M, Sánchez L, et al (2015) Genomic selection prediction accuracy in a
616 perennial crop: case study of oil palm (Elaeis guineensis Jacq.). Theor Appl Genet
617 128:397–410

De Franqueville H, Renard JL (1990) Improvement of oil palm vascular wilt tolerance-results and development of the disease at the R. Michaux plantation. Oleagineux Vol.45 No.10 pp.399-405.

Durand-Gasselin T, Asmady H, Flori A, et al (2005) Possible sources of genetic resistance in oil palm (Elaeis guineensis Jacq.) to basal stem rot caused by Ganoderma boninense– prospects for future breeding. Mycopathologia 159:93–100

Durand-Gasselin T, Cochard B, de Franqueville H (2018) Advances in disease-resistant oil palm varieties. In: Center for International Cooperation in Agricultural Research for Development (CIRAD), France, Rival A (eds) Burleigh Dodds Series in Agricultural Science. Burleigh Dodds Science Publishing, pp 137–164

Faizah R, Putranto RA, Wening S, et al (2020) Differential expression of root specific genes of oil palm seedlings at early stage of *Ganoderma boninense* infection. IOP Conf Ser Earth Environ Sci 418:012044. https://doi.org/10.1088/1755-1315/418/1/012044

Franqueville H de, Asmady H, Jacquemard JC, et al (2001) Indications on sources of oil palm (Elaeis guineensis Jacq.) genetic resistance and susceptibility to Ganoderma sp., the cause of basal stem rot. In: Cutting-edge technologies for sustained competitiveness: Proceedings of the 2001 PIPOC International Palm Oil Congress, Agriculture Conference, Kuala Lumpur, Malaysia, 20-22 August 2001. Malaysian Palm Oil Board (MPOB), pp 420–431

Garin V, Wimmer V, Mezmouk S, et al (2017) How do the type of QTL effect and the form of the residual term influence QTL detection in multi-parent populations? A case study in the maize EU-NAM population. Theor Appl Genet 130:1753–1764. https://doi.org/10.1007/s00122-017-2923-3

Gascon JP, De Berchoux CH (1964) Caractéristiques de la production d'Elaeis guineensis (Jacq.) de diverses origines et leurs croisements. Appl À Sélection Palmier À Huile Oléagineux 19:75–84

George AW, Visscher PM, Haley CS (2000) Mapping quantitative trait loci in complex pedigrees: a two-step variance component approach. Genetics 156:2081–2092

Govender N, Abu-Seman I, Mui-Yun W (2020) Root Lignin Composition and Content in Oil Palm (Elaeis guineensis Jacq.) Genotypes with Different Defense Responses to Ganoderma boninense. Agronomy 10:1487. https://doi.org/10.3390/agronomy10101487

Govender NT, Mahmood M, Seman IA, Wong M-Y (2017) The Phenylpropanoid Pathway and Lignin in Defense against Ganoderma boninense Colonized Root Tissues in Oil Palm (Elaeis guineensis Jacq.). Front Plant Sci 8:1395

Green P, Falls K, Crooks S (1990) CRIMAP Documentation. https://www.animalgenome.org/hu/CRIMAPwkshp/crimap-doc.html. Accessed 5 Apr 2018

Hama-Ali EO, Panandam JM, Tan SG, et al (2015) Association between basal stem rot disease and simple sequence repeat markers in oil palm, Elaeis guineensis Jacq. Euphytica 202:199–206

659 Ho C-L, Tan Y-C, Yeoh K-A, et al (2016) De novo transcriptome analyses of host-fungal
660 interactions in oil palm (Elaeis guineensis Jacq.). BMC Genomics 17(1):1-19.

661 Idris A, Kushairi A, Ismail S, Ariffin D (2004) Selection for partial resistance in oil palm
662 progenies to Ganoderma basal stem rot. J Oil Palm Res 16:12–18

663 Jeennor S, Volkaert H (2014) Mapping of quantitative trait loci (QTLs) for oil yield using
664 SSRs and gene-based markers in African oil palm (Elaeis guineensis Jacq.). Tree
665 Genet Genomes 10:1–14

666 Johnson M, Kumar A, Oladzad-Abbasabadi A, et al (2019) Association Mapping for 24 Traits
667 Related to Protein Content, Gluten Strength, Color, Cooking, and Milling Quality
668 Using Balanced and Unbalanced Data in Durum Wheat [Triticum turgidum L. var.
669 durum (Desf).]. Front Genet 10:. https://doi.org/10.3389/fgene.2019.00717

670 Korontzis G, Malosetti M, Zheng C, et al (2020) QTL detection in a pedigreed breeding
671 population of diploid potato. Euphytica 216(9):1-14. https://doi.org/10.1007/s10681-
672 020-02674-y

673 Lee M, Xia JH, Zou Z, et al (2015) A consensus linkage map of oil palm and a major QTL for
674 stem height. Sci Rep 5:(1), 1-7.

675 Mangandi J, Verma S, Osorio L, et al (2017) Pedigree-based analysis in a multiparental
676 population of octoploid strawberry reveals QTL alleles conferring resistance to
677 Phytophthora cactorum. G3 Genes Genomes Genet 7:1707–1719

678 Mercière M, Laybats A, Carasco-Lacombe C, et al (2015) Identification and development of
679 new polymorphic microsatellite markers using genome assembly for Ganoderma
680 boninense, causal agent of oil palm basal stem rot disease. Mycol Prog 14:103

681 Meunier J, Gascon JP (1972) Le schéma général d'amélioration du palmier à huile à l'IRHO.
682 Oléagineux 27:1–12

683 Muranty H, Jorge V, Bastien C, et al (2014) Potential for marker-assisted selection for forest
684 tree breeding: lessons from 20 years of MAS in crops. Tree Genet Genomes 10:1491–
685 1510

686 Nelson R, Wiesner-Hanks T, Wisser R, Balint-Kurti P (2018) Navigating complexity to breed
687 disease-resistant crops. Nat Rev Genet 19:21–33. https://doi.org/10.1038/nrg.2017.82

688 Nusaibah SA, Akmar ASN, Idris AS, et al (2016) Involvement of metabolites in early defense
689 mechanism of oil palm (Elaeis guineensis Jacq.) against Ganoderma disease. Plant
690 Physiol Biochem 109:156–165

691 Nyouma A, Bell JM, Jacob F, Cros D (2019) From mass selection to genomic selection: one
692 century of breeding for quantitative yield components of oil palm (Elaeis guineensis
693 Jacq.). Tree Genet Genomes 15(5):1-16. https://doi.org/10.1007/s11295-019-1373-2

694 Paterson RRM (2019) Ganoderma boninense Disease of Oil Palm to Significantly Reduce
695 Production After 2050 in Sumatra if Projected Climate Change Occurs.
696 Microorganisms 7:24

Paterson RRM (2007) Ganoderma disease of oil palm—A white rot perspective necessary for integrated control. Crop Prot 26:1369–1376. https://doi.org/10.1016/j.cropro.2006.11.009

Poland J, Rutkoski J (2016) Advances and Challenges in Genomic Selection for Disease Resistance. Annu Rev Phytopathol 54:79–98. https://doi.org/10.1146/annurev-phyto-080615-100056

Poland JA, Balint-Kurti PJ, Wisser RJ, et al (2009) Shades of gray: the world of quantitative disease resistance. Trends Plant Sci 14:21–29

Pootakham W, Jomchai N, Ruang-areerate P, et al (2015) Genome-wide SNP discovery and identification of QTL associated with agronomic traits in oil palm using genotyping-by-sequencing (GBS). Genomics 105:288–295

Rees RW, Flood J, Hasan Y, et al (2009) Basal stem rot of oil palm (Elaeis guineensis); mode of root infection and lower stem invasion by Ganoderma boninense. Plant Pathol 58:982–989

Rees RW, Flood J, Hasan Y, Cooper RM (2007) Effects of inoculum potential, shading and soil temperature on root infection of oil palm seedlings by the basal stem rot pathogen Ganoderma boninense. Plant Pathol 56:862–870. https://doi.org/10.1111/j.1365-3059.2007.01621.x

Rochmyaningsih D (2019) Making peace with oil palm. Science 365:112–115. https://doi.org/10.1126/science.365.6449.112

Roux F, Voisin D, Badet T, et al (2014) Resistance to phytopathogens *e tutti quanti* : placing plant quantitative disease resistance on the map: Quantitative disease resistance in plants. Mol Plant Pathol 15:427–432. https://doi.org/10.1111/mpp.12138

Sakeh NM, Abdullah SNA, Bahari MNA, et al (2020) EgJUB1 and EgERF113 transcription factors as master regulators of defense response in Elaeis guineensis against the hemibiotrophic *Ganoderma boninense*. BMC plant biology, 21(1), 1-20.

Singh R, Low E-TL, Ooi LC-L, et al (2013) The oil palm SHELL gene controls oil yield and encodes a homologue of SEEDSTICK. Nature 500:340–344. https://doi.org/10.1038/nature12356

Smith AB, Cullis BR, Thompson R (2005) The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. J Agric Sci 143:449–462. https://doi.org/10.1017/S0021859605005587

Tan Y-C, Yeoh K-A, Wong M-Y, Ho C-L (2013) Expression profiles of putative defence-related proteins in oil palm (Elaeis guineensis) colonized by Ganoderma boninense. J Plant Physiol 170:1455–1460. https://doi.org/10.1016/j.jplph.2013.05.009

Team RC (2012) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2012. ISBN 3-900051-07-0

Tee S-S, Tan Y-C, Abdullah F, et al (2013) Transcriptome of oil palm (Elaeis guineensis Jacq.) roots treated with Ganoderma boninense. Tree Genet Genomes 9:377–386

Tisné S, Denis M, Cros D, et al (2015) Mixed model approach for IBD-based QTL mapping in a complex oil palm pedigree. BMC Genomics 16(1):1-12.

Tisné S, Maurin G, Bink M, et al (2019) Complex Trait Improvement in the Reciprocal Recurrent Selection Context using a Pedigree Based QTL Mapping Approach. In: Proceedings of the PIPOC 2019 International Palm Oil Congress Agriculture, Biotechnology & Sustainability Conference. Malaysian Palm Oil Board (MPOB), Kuala Lumpur Convention Centre, Kuala Lumpur, Malaysia, pp 356–362

Tisné S, Pomiès V, Riou V, et al (2017) Identification of Ganoderma disease resistance loci using natural field infection of an oil palm multiparental population. G3 Genes Genomes Genet 7:1683–1692

Tranbarger TJ, Kluabmongkol W, Sangsrakru D, et al (2012) SSR markers in transcripts of genes linked to post-transcriptional and transcriptional regulatory functions during vegetative and reproductive development of Elaeis guineensis. BMC Plant Biol 12(1):1-12. https://doi.org/10.1186/1471-2229-12-1

van de Weg E, Di Guardo M, Jänsch M, et al (2018) Epistatic fire blight resistance QTL alleles in the apple cultivar 'Enterprise' and selection X-6398 discovered and characterized through pedigree-informed analysis. Mol Breed 38(1):1-18. https://doi.org/10.1007/s11032-017-0755-0

van de Weg WE, Voorrips RE, Finkers R, et al (2004) Pedigree genotyping: a new pedigree-based approach of QTL identification and allele mining. In Acta Hortic 45–50. https://doi.org/10.17660/ActaHortic.2004.663.1

van Eeuwijk FA, Boer M, Totir LR, et al (2010) Mixed model approaches for the identification of QTLs within a maize hybrid breeding program. Theor Appl Genet 120:429–440

Verma S, Whitaker VM (2018) Prediction of QTL genotypes and trait phenotypes using FlexQTLTM: a pedigree-based analysis approach. J. Plant Biol. Crop Res, 2, 1006.

Voorrips RE (2002) MapChart: Software for the Graphical Presentation of Linkage Maps and QTLs. J Hered 93:77–78. https://doi.org/10.1093/jhered/93.1.77

Woittiez LS, van Wijk MT, Slingerland M, et al (2017) Yield gaps in oil palm: A quantitative review of contributing factors. Eur J Agron 83:57–77

Würschum T (2012) Mapping QTL for agronomic traits in breeding populations. Theor Appl Genet 125:201–210

Xue S, Ogut F, Miller Z, et al (2017) Comparison of one-stage and two-stage genome-wide association studies. bioRxiv 099291. https://doi.org/10.1101/099291

Yeoh K-A, Othman A, Meon S, et al (2013) Sequence analysis and gene expression of putative oil palm chitinase and chitinase-like proteins in response to colonization of Ganoderma boninense and Trichoderma harzianum. Mol Biol Rep 40:147–158. https://doi.org/10.1007/s11033-012-2043-8

Yeoh K-A, Othman A, Meon S, et al (2012) Sequence analysis and gene expression of putative exo- and endo-glucanases from oil palm (Elaeis guineensis) during fungal infection. J Plant Physiol 169:1565–1570. https://doi.org/10.1016/j.jplph.2012.07.006

Zaki NM, Singh R, Rosli R, Ismail I (2012) Elaeis oleifera Genomic-SSR Markers: Exploitation in Oil Palm Germplasm Diversity and Cross-Amplification in Arecaceae. Int J Mol Sci 13:4069–4088. https://doi.org/10.3390/ijms13044069

Zhang Z, Ober U, Erbe M, et al (2014) Improving the accuracy of whole genome prediction for complex traits using the results of genome wide association studies. PloS One 9:e93017

Figure captions

Fig. 1 Pedigree of the pre-nursery GB oil palm population. Boxes on the left represent the founders of the La Mé (LM, panel A) and Yangambi (YBI, panel B) populations. Note that the La Mé founders LM_1:4 are full sibs. Other boxes represent full-sib families whose color represents their relation to their genetic background, with the number of individuals in parenthesis. The circled cross symbols represent progenies obtained through self-pollination, and successive self-pollinated progenies keep the same color.

Fig. 2 Distribution of *Ganoderma* disease resistance in the pre-nursery GB oil palm population. Distribution of the percentage of affected individuals in crosses (A), BLUP obtained from random effect of the GCA in GB in a GLMM (B) and LMM (C) for the La Mé (LM) and Yangambi (YBI) populations. Different colors represent different genetic backgrounds.

Fig. 3 QTL mapping of *Ganoderma* resistance in the pre-nursery GB oil palm population. QTL regions marked by FlexQTL software in six independent simulations (LMM and GLMM models, three random starting seeds) (A) and the averaged posterior intensity calculated at a 1 cM grid for the six simulations (B) are plotted along the genome. In panel A, the yellow to red color code scale depict the value of intensity of the corresponding marked QTL regions found in the "MQTRegions.new" FlexQTL output file. In panel B, a white to red color scale indicates the number of marked QTL regions among the six simulations at the corresponding position in the genome.

Fig. 4 Variance components of *Ganoderma* resistance in the pre-nursery screening tests. Variance components are plotted as a percentage of the total phenotypic variance for each of the steps performed in the stepwise selection model. GA/GB: heterotic group A and B; GCA: general combining ability; SCA: Specific combining ability; CL: *Ganoderma* clonal lines; QTL names: see M&M section.

Fig. 5 Pre-nursery QTL effects on *Ganoderma* resistance to natural field infection in the La Mé genetic background. (A) Survival curves of the La Mé population according to the percentage of favorable alleles at the 21 La Mé QTL detected in the pre-nursery analysis, the red to green color scale indicates an increasing percentage. Survival estimates are plotted at the time of the first observation of a *Ganoderma* symptom. (B) Scatterplot showing the relationship between the statistical significances of QTL effects in the pre-nursery experiments (posterior intensity, x-axis) and in the field (-log (P-value) from the Cox model, y-axis). QTL originate from QTL mapping using the GB (squares) or LM (triangles) pedigree. Consistency between field and pre-nursery QTL effects was defined for QTL alleles decreasing the number of affected progenies in the pre-nursery trials and delaying the appearance of the first symptom of *Ganoderma*: inconsistent and consistent QTL effects are depicted by green (+) or red (-) symbols, respectively. QTL for which one of the three allelic classes (*QQ*, *Qq* or *qq*) was represented by less than ten individuals are depicted by shaded symbols. QTL names: see M&M section.

Figure 3

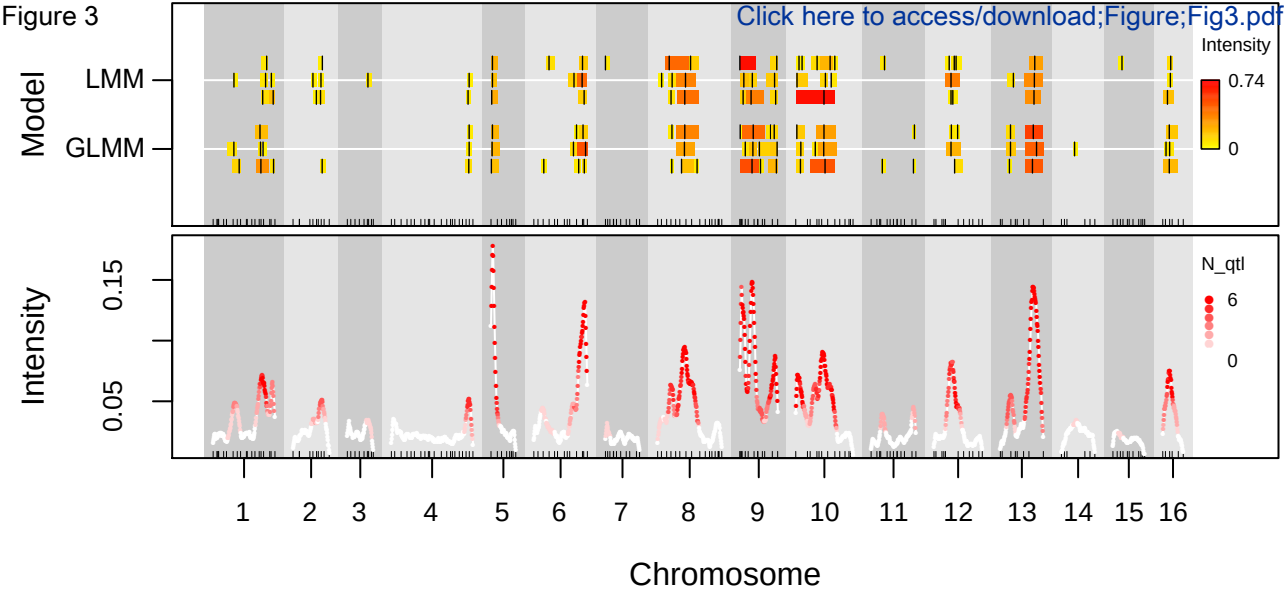


Figure 4

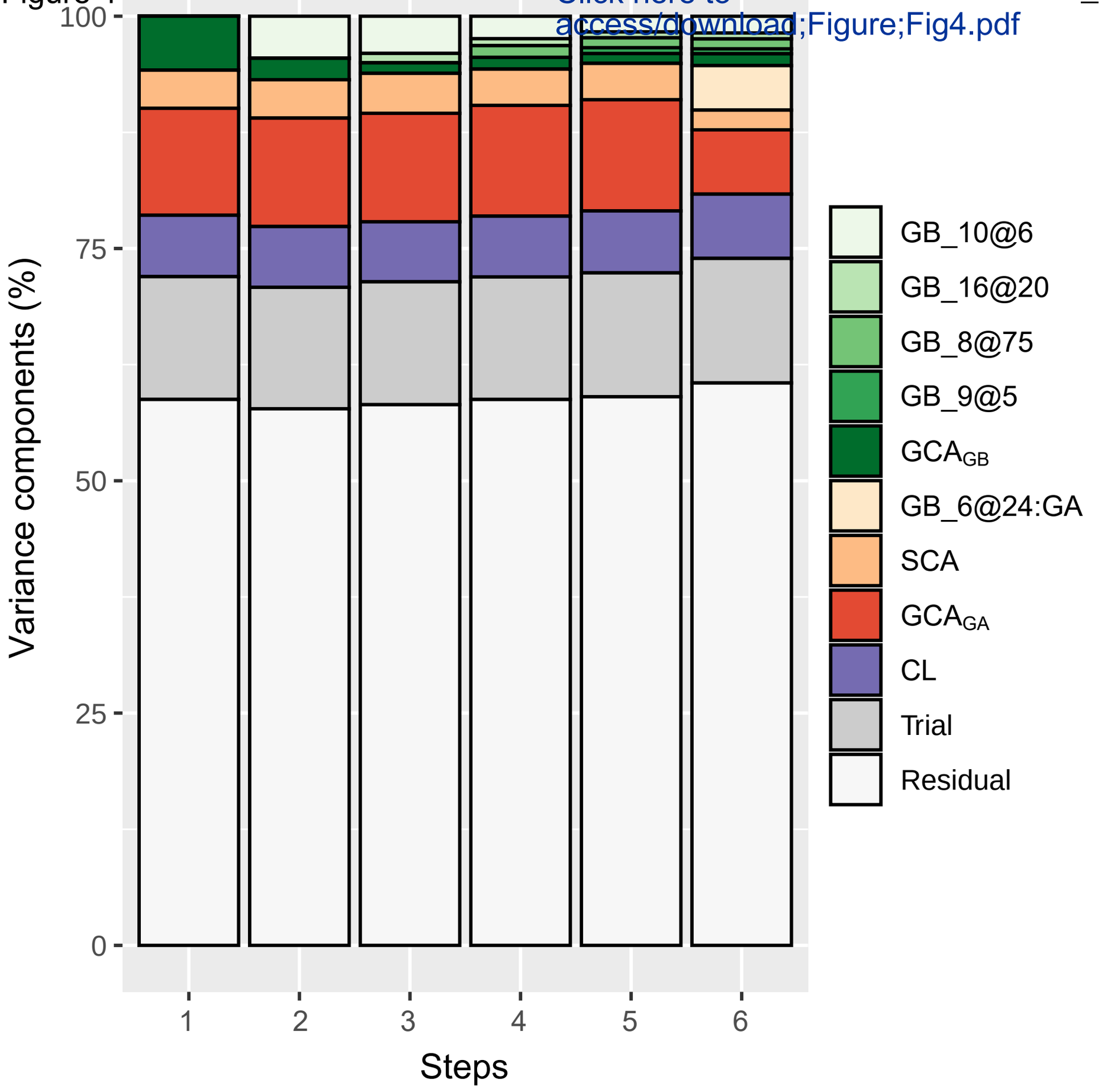


Figure 5

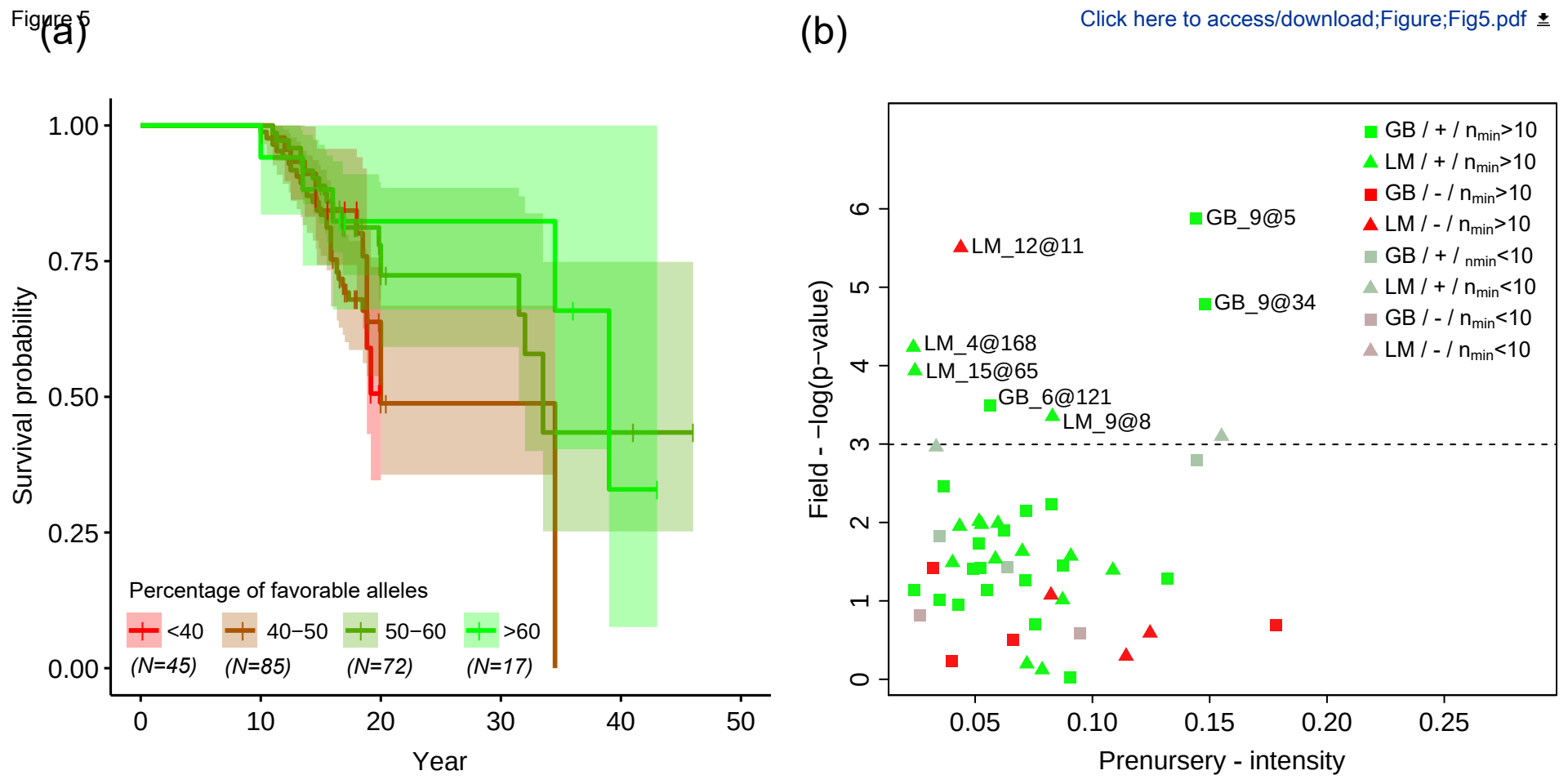


Figure 2

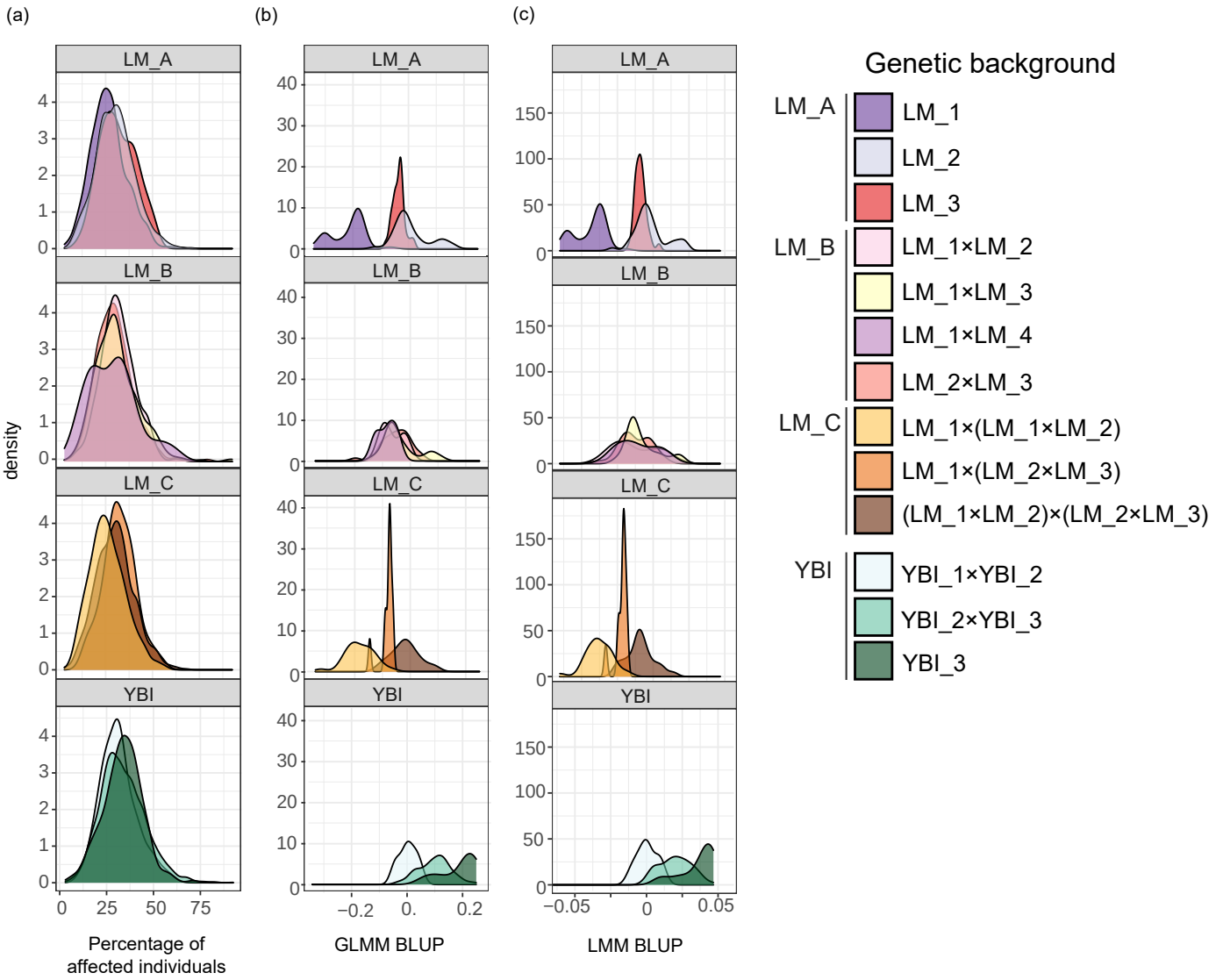
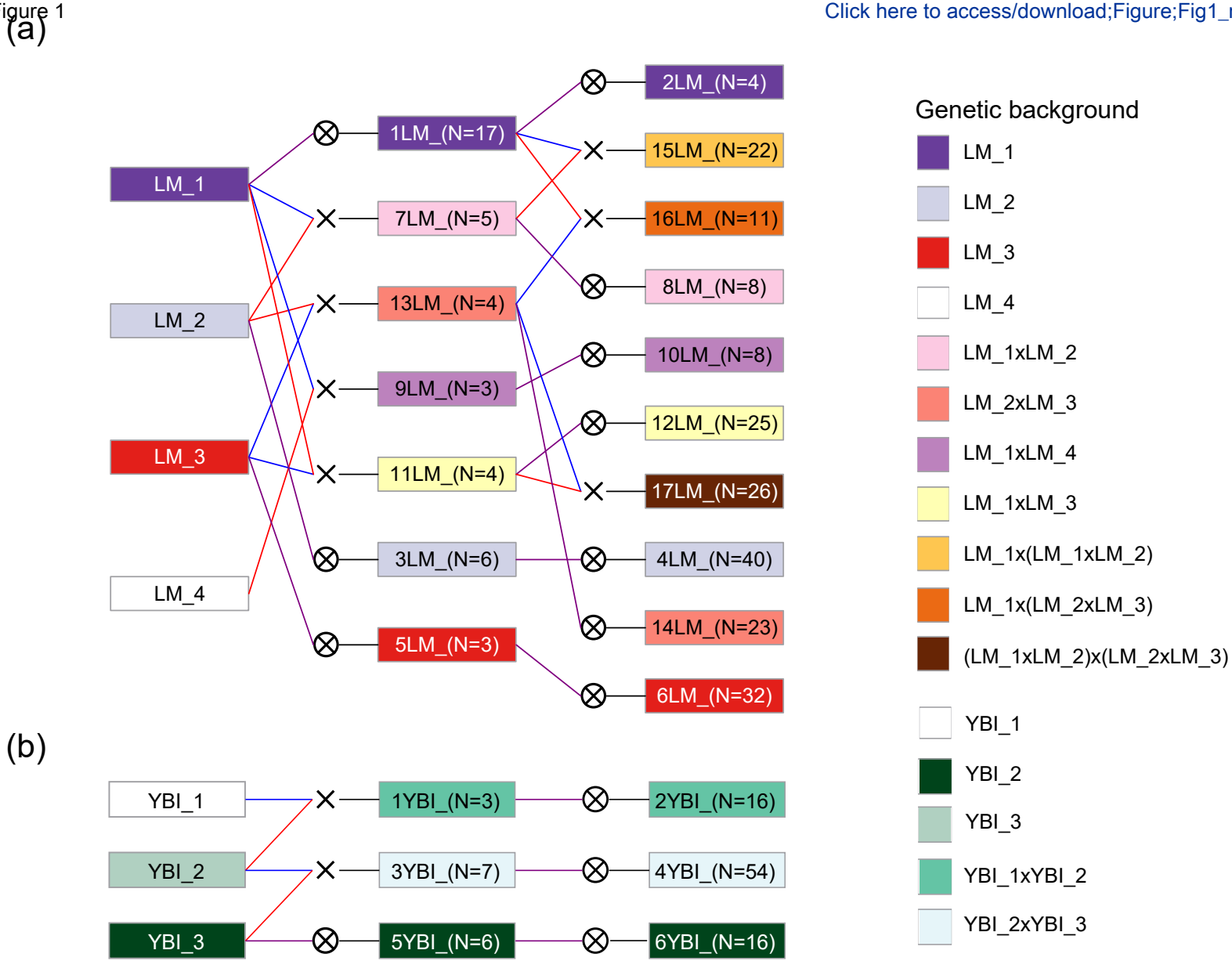


Figure 1

Click here to access/download;Figure;Fig1_revised.pdf



Title Page

Title

In silico QTL mapping in an oil palm breeding program reveals a quantitative and complex genetic resistance to *Ganoderma boninense*

Authors

Aurélie Daval, Virgine Pomiès, Sandrine le Squin, Marie Denis, Virginie Riou, Frederic Breton, Nopariansyah, Marco Bink, Benoît Cochard, Florence Jacob, Norbert Billotte and Sébastien Tisné

Author information

Aurélie Daval : CIRAD, UMR AGAP, F-34398 Montpellier, France ; AGAP, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France ; ORCID ID: 0000-0002-2613-0562

Virgine Pomiès : CIRAD, UMR AGAP, F-34398 Montpellier, France ; AGAP, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France ; ORCID ID: 0000-0002-5481-5120

Sandrine le Squin : PalmElit SAS, Montferrier-sur-Lez, France

Marie Denis : CIRAD, UMR AGAP, F-34398 Montpellier, France ; AGAP, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France ; ORCID ID : 0000-0002-1693-9894

Virginie Riou : CIRAD, UMR AGAP, F-34398 Montpellier, France ; AGAP, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France

Frederic Breton : CIRAD, UMR AGAP, F-34398 Montpellier, France ; AGAP, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France ; ORCID ID: 0000-0002-6853-2623

Nopariansyah: P.T SOCFINDO, Jl. Yos Sudarso, Medan, Sumatera Utara 20115, Indonesia

Marco Bink: Biometris, Wageningen UR, PO Box 16, 6700 AA Wageningen, The Netherlands; Current address: Research & Technology Center, Hendrix Genetics, Boxmeer, The Netherlands ; ORCID ID: 0000-0002-1278-2092

Benoît Cochard : PalmElit SAS, Montferrier-sur-Lez, France

Florence Jacob : PalmElit SAS, Montferrier-sur-Lez, France ; ORCID ID : 0000-0002-0454-1037

Norbert Billotte : CIRAD, UMR AGAP, F-34398 Montpellier, France ; AGAP, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France ; ORCID ID : 0000-0002-0438-0966

Sébastien Tisné (Corresponding author): CIRAD, UMR AGAP, F-34398 Montpellier, France ; AGAP, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France; E-mail address: sebastien.tisne@cirad.fr; ORCID ID: 0000-0001-9838-3975

Abstract

Basal stem rot caused by *Ganoderma boninense* is the major threat to oil palm cultivation in South-East Asia, which accounts for 80% of palm oil production worldwide, and this disease is increasing in Africa. The use of resistant planting material as part of an integrated pest management of this disease is one sustainable solution. However, breeding for *Ganoderma* resistance requires long-term and costly research, which could greatly benefit from marker assisted selection (MAS). In this study, we evaluated the effectiveness of an *in silico* genetic mapping approach that took advantage of extensive data recorded in an ongoing breeding program. A pedigree-based QTL mapping approach applied to more than 10 years' worth of data collected during pre-nursery tests revealed the quantitative nature of *Ganoderma* resistance and identified underlying loci segregating in genetic diversity that is directly relevant for the breeding program supporting the study. To assess the consistency of QTL effects between pre-nursery and field environments, information was collected on the disease status of the genitors planted in genealogical gardens and modeled with pre-nursery-based QTL genotypes. In the field, individuals were less likely to be infected with *Ganoderma* when they carried more favorable alleles at the pre-nursery QTL. Our results pave the way for a MAS of *Ganoderma* resistant and high yielding planting material and the provided proof-of-concept of this efficient and cost-effective approach could motivate similar studies based on diverse breeding programs.

Keywords

Oil palm, basal stem rot disease, nursery screening test, breeding population, QTL mapping, pedigree-based analysis.

Acknowledgments

This study was based on a very intensive and laborious work involving many people in the long-term. We thank Zulkifi Lubis, Augustiaman Purba, Shri Jeweyen, and all the SOCFIN Indonesia staff at Tanah Gambus who

performed the pre-nursery trials. We thank the PalmElit staff, Hubert de Franqueville and Michaël Pernaci for information on plant pathology and Nicolas Turnbull on breeding. We acknowledge Tristan Durand-Gasselin (PalmElit) for his insightful review of the study and manuscript. We thank Eric van de Weg (Wageningen UR) for the review of the manuscript.

This research was partly funded by a grant from PalmElit SAS. MD contributed partly to this study while she was visiting researcher at Georgetown University and supported by the European Union's Horizon 2020 research and innovation program under grant agreement No840383.

CRediT authorship contribution statement

Conceptualization: Benoît Cochard, Sébastien Tisné ; Data Curation: Sandrine le Squin, Virginie Riou, Florence Jacob ; Formal analysis: Aurélie Daval, Sandrine le Squin, Sébastien Tisné; Investigation: Virgine Pomiès, Frederic Breton, Nopariansyah; Methodology: Sandrine le Squin, Marie Denis, Sébastien Tisné; Project administration: Florence Jacob, Sébastien Tisné ; Resources: Nopariansyah, Benoît Cochard ; Software: Marco Bink ; Supervision: Florence Jacob, Norbert Billotte, Sébastien Tisné ; Writing - Original Draft: Aurélie Daval, Sébastien Tisné ; Writing - Review & Editing: Aurélie Daval, Marie Denis, Frederic Breton, Benoît Cochard, Florence Jacob, Norbert Billotte, Sébastien Tisné ; Visualization: Aurélie Daval, Sébastien Tisné.

1. Introduction

The African oil palm (*Elaeis Guineensis* Jacq.) is the leading oil crop worldwide with a global annual production of around 73 Mt, and accounts for more than 35% of all the edible vegetable oil produced worldwide (USDA statistics, 2019). Oil palm is expected to be able to respond to the global increase in the demand for vegetable oil projected to be 240 Mt in 2050, even higher if its non-food uses are included (Corley 2009). The oil palm sector has agreed on sustainability goals to reach this global demand (Rochmyaningsih 2019), in particular through the certification of sustainable produced palm oil (the Roundtable on Sustainable Palm Oil, RSPO, https://rspo.org/). However, pests and diseases threaten palm oil production in all areas of cultivation and contribute to the current yield gap (Woittiez et al. 2017). If it is to achieve the zero-deforestation goal in high conservation value forests included in the RSPO commitments, oil palm will inevitably be cultivated on existing arable lands under increasing pathogen pressure. The integrated pest management (IPM) covers sustainable solutions to this problem including improved plant disease resistance. Oil palm breeders thus needs to focus on developing resistant planting material, while maintaining or even improving oil yield.

The basal stem rot disease caused by *Ganoderma boninense* is a major threat in South East Asia, with projections worsening due to climate change (Paterson 2019). This pathogenic fungus is a soil-borne basidiomycete that mainly infects the oil palm when its roots come into contact with infected debris or with the roots of neighbor palms (Rees et al. 2009). *Ganoderma* stem rot disease has a significant effect on oil yield even when only 10-20% of palm trees are infected, and 30-70% of the trees may have died over a typical 25-year planting cycle (Durand-Gasselin et al. 2005; Cooper et al. 2011). To date, no specific interaction and/or complete resistance have been identified in oil palm/*Ganoderma* pathosystem, which is consistent with its hemibiotrophic pathogenic lifestyle. However, observations of contrasted levels of resistance in diverse genetic backgrounds suggest that breeding for quantitative disease resistance (QDR) is a promising solution (Franqueville et al. 2001). Typically, research on perennial plant disease resistance is based on large scale costly field experiments, even more so when investigating QDR. When possible, *ex situ* experiments with controlled inoculation of the pathogen are powerful tools that offer more repeatability and increase both speed and throughput, especially in genetic surveys. In oil palm, such pre-nursery tests were first developed for research on vascular wilt (De Franqueville and Renard 1990), followed by *Ganoderma* in the 2000s (Idris et al. 2004; Breton et al. 2006b; Rees et al. 2007) and are now widely used. However, transferring results to the field can be problematic because of a more complex biotic context, the age specificity of the QDR mechanisms, or the effects of cultural practice management on disease epidemiology. Despite these challenges, by combining field and pre-nursery approaches in long-term works in the framework of an oil palm breeding program, Cirad, its subsidiary PalmElit, and their partners have managed to release planting material that is highly resistant to vascular wilt and intermediate resistant to basal stem rot caused by *Ganoderma* (De Franqueville and Renard 1990; Franqueville et al. 2001; Durand-Gasselin et al. 2005; Breton et al. 2009).

Information on the genetic architecture and molecular determinisms of traits of interest could help shorten the long breeding cycle of oil palm, which currently exceeds 20 years, and would be particularly useful in the case of *Ganoderma* disease given the cumbersome nature of field and nursery trials. Marker assisted selection (MAS) based on this information would increase the annual genetic gain thanks to both accelerated evaluation of selection candidates and increased selection intensity by enabling surveys of wider genetic diversity at the same cost (Cros et al. 2015, 2017). Moreover, identification of the genetic bases of resistance to *Ganoderma* could resolve the challenge of breeding for both QDR and yield related traits (Nelson et al. 2018) by using simulation and prediction tools (Tisné et al. 2019). Most molecular studies on *Ganoderma* disease to date have been based on inoculated *vs* non-inoculated seedlings at the pre-nursery stage, with no or low genetic diversity. The first investigations focused on *a priori* selection of candidate resistance genes to fungal diseases (Yeoh et al. 2012, 2013; Tan et al. 2013).

Next the genes, proteins and pathways affected by *Ganoderma* infection were identified using broader transcriptomic (Tee et al. 2013; Ho et al. 2016; Bahari et al. 2018; Faizah et al. 2020; Sakeh et al. 2020), proteomic (Al-Obaidi et al. 2014) and metabolomic (Nusaibah et al. 2016) approaches. Considering that *Ganoderma* is a white rot fungus (Paterson 2007), lignin related traits were investigated as putative QDR mechanisms by surveying the response of lignin content and composition to *Ganoderma* infection together with the associated genes (Govender et al. 2017). Lignin related traits and nutritional traits were found to differ in progenies with different levels of resistance to *Ganoderma* (Govender et al. 2020) but the restrained genetic design confounds the effects of genetic and resistance variation.

QTL mapping offers an alternative approach that provides information on the genetic architecture based on a relevant genetic diversity, with no *a priori* biological knowledge. The detected loci form the basis of the MAS strategy but also provide insights into the mechanisms and genes involved in the QDR. The first published QTL study reported the analysis of 79 individuals from one resistant and two susceptible families based on 58 simple sequence repeat markers and found alleles associated with *Ganoderma* symptoms (Hama-Ali et al. 2015). More conclusive insights would require much more data, but QTL analyses of oil palm crosses are typically not sufficiently effective due to biological and cost constraints (Jeennor and Volkaert 2014; Lee et al. 2015; Pootakham et al. 2015). This is even more problematic for field studies that are indispensable to assess genetic diversity in an agronomic context, whose implementation is very costly and would result in lower production income due to the disease context. A powerful and cost-effective approach is to directly use the databases compiled in ongoing breeding programs, which are typically large and obtained from diverse relevant genetic backgrounds, to map *in silico* the QTLs for the traits of interest (Parisseaux and Bernardo 2004). Despite the potential of this approach, data from breeding programs are unique, mainly because of a complex genetic design that may be biased due to selection, or unbalanced phenotyping coverage. Thus, they require appropriate statistical models for their development and evaluation in contrasted contexts, which are currently an active research topic (Würschum 2012; Garin et al. 2017; Korontzis et al. 2020). In oil palm, an *in silico* QTL mapping approach based on the two step variance component approach considering identity by descent (IBD) information (George et al. 2000; van Eeuwijk et al. 2010) yielded promising results on production traits recorded in large scale evaluation genetic trials (Tisné et al. 2015). This approach was successfully extended to survival data and applied to a multi-parent population to detect *Ganoderma* resistance QTLs in the field, allowing to identify two QTL related to the occurrence of the first disease symptoms, and two related to the death due to *Ganoderma* (Tisné et al. 2017). A Bayesian approach to pedigree based QTL mapping using IBD information was also developed in the 2000s and implemented in the

FlexQTL software (van de Weg et al. 2004; Bink et al. 2008). This made it possible to carry out increasing numbers of studies in several crops that share the constraints and potential described above for oil palm, in particular for disease resistance in strawberry (Mangandi et al. 2017; Anciro et al. 2018) or in apple (van de Weg et al. 2018). In this study, we evaluated the potential of an *in silico* approach based on the large existing databases of a long-term oil palm breeding program for the study of *Ganoderma* resistance. We genotyped an existing DNA bank primarily established for identity checking purpose and performed a pedigree-based QTL mapping using data recorded in *Ganoderma* pre-nursery trials over a period of more than ten years. We then assessed the consistency of pre-nursery QTL effects in natural field conditions using a database recording the *Ganoderma* infection status over years for the palms planted in genealogical gardens. Thus, using a cost-effective approach that is directly relevant to the breeding program, we were able to study two major issues, i.e. the genetic architecture and consistency between pre-nursery and field results, paving the way for the implementation of MAS for *Ganoderma* resistant planting material.

2. Material and methods

2.1. Plant material

The palm trees used in this study belong to the oil palm breeding program of Cirad, its subsidiary PalmElit and their partner PT Socfin Indonesia (Indonesia). This breeding program is conducted in a recurrent reciprocal selection scheme with two heterotic groups A and B (GA and GB to produce superior GA×GB hybrid crosses used as commercial planting material (Gascon and De Berchoux 1964; Meunier and Gascon 1972). Individuals from different heterotic groups have complementary yield component traits, with low fruit bunch number and high bunch weight in GA and reciprocally in GB. GA×GB hybrids consequently show a heterosis effect on fruit bunch yield. Moreover, individuals included in GA are Dura palms, homozygous for the thick alleles of the shell gene (Singh et al. 2013) while individuals included in GB are Pisifera (homozygous alternative alleles), the hybrid GA×GB being Tenera which is the most productive form with thin shell. The parental population studied for the *Ganoderma* resistance included only individuals from GB, grouping genetic origins of La Mé (LM, Ivory Coast) and Yangambi (YBI, Republic Democratic of Congo). The GB pedigree used in the pre-nursery analysis comprised 372 individuals including founders, with 246/126 from LM/YBI genetic origin respectively and 240 /93 genotyped (Supp. Table 1). Among them 200 LM and 83 YBI parents were directly progeny tested for *Ganoderma* resistance in a pre-nursery screening test (Fig. 1). The individuals were distributed over many full-sib families derived from a small number of founders through consecutive crosses or self-pollinations in the framework of the ongoing

breeding program (Fig. 1). Among the 372 individuals in the whole pedigree, 219 LM individuals were planted between 1970s and 2000s at the same location (Bangun Bandar, Indonesia) and were used for subsequent field analysis.

2.2. Phenotypic data

2.2.1. Pre-nursery screening tests

An early pre-nursery screening test was developed in the 2000s by Cirad and Socfin Indonesia in the Tanah Gambus estate, Indonesia. The first objective was to speed up the evaluation of genetic resistance to *Ganoderma* of commercial oil palm planting materiel, using controlled and standardized inoculation of germinated seeds (Breton et al., 2006). The inoculation of germinated seed was performed using a 12 week-old *Ganoderma*-colonized rubber wood block (108 cm3) as inoculum source, that was previously deposited in the nursery polybag before the seeds were planted.

A pure dikariotic *Ganoderma boninense* isolate was used in all the trials (NJ), previously harvested from an infected oil palm planted in Bangun Bandar, SOCFINDO estate (Mercière et al. 2015). This isolate was successively regenerated from the bole of young infected seedlings in consecutive pre-nursery trials to provide several dikariotic clonal lines (CL, n=7) over the 10 years of testing. These reactivating steps of the isolate made it possible to avoid the loss of pathogenicity often observed after successive sub-cultures on artificial fungi growth media (Butt et al. 2006). A single pathogen CL was used for all the crosses tested in a single trial. Around 100 crosses were assessed simultaneously in each pre-nursery trial. Among them, 20% were control crosses from susceptible, intermediate and resistant genetic backgrounds and were included in all the trials performed. Of the remaining 80% of crosses representing the tested crosses, 50% overlapped two consecutive trials, leading to at least two independent tests per tested cross. Each cross was represented by 100 inoculated germinated seeds clustered in five replicates following the protocol described by Breton et al. (2009). Inoculated seedlings were observed every four weeks for the appearance of the first external disease symptom, on average between 8 and 12 weeks after inoculation of the germinated seeds, after which the disease symptoms were recorded at two weekly intervals as (1) infected and (0) if not infected. The trial was stopped when the average percentage of infected seedlings within the group of control crosses reached 30%, usually around 34 weeks after inoculation of the germinated seeds. This 30% threshold was determined to have the best “discriminating power” between the resistant and sensitive control crosses, and so among the tested progenies (Breton et al. 2009).

This study included 102 *Ganoderma* pre-nursery screening test trial, covering 10 years of data recording. The trials performed between 2007 and 2017 represented the evaluation of 4,017 unique crosses, from either GA×GA, GA×GB or GB×GB genetic background. Considering that the purpose of this study was to assess the genetic bases of *Ganoderma* resistance in the commercial genetic material, only the GA×GB crosses were taken into consideration (n=3,792), derived from 2,037 and 340 individuals from the GA and GB respectively. Each parent from GB included in the analysis was progeny tested in an average of 20.5 GA×GB crosses.

2.2.2. Statistical modeling of pre-nursery data

The resistance of the GB individuals was progeny-tested through several GA×GB crosses involving them as GB parents. The response variable Y considered in this study was the proportion of affected progenies per cross at the end of the trial. A first step of statistical modeling of Y was necessary to obtain a single value per genotype required for the QTL analysis while accounting for nuisance effects due to the long-term data. Y was modeled using generalized linear mixed models (GLMM). Briefly, in a GLMM, Y is assumed to be generated by a particular distribution in the exponential family. The conditional mean of the distribution μ is linked to a linear predictor η which contains fixed and random effects, through the inverse link function g^{-1}:

$$g(\mu) = \eta = X\beta + Z_T u_T + Z_A u_A + Z_B u_B + Z_C u_C$$

where X is a $n \times m$ design matrix relating observations to *Ganoderma boninense* CL fixed effects β where β is a $m \times 1$ vector ($m = 7$), Z_T is a $n \times t$ design matrix relating observations to trial random effects $u \sim N(0, I\sigma_T^2)$ with u is a $t \times 1$ vector ($t = 102$), Z_C is a $n \times c$ design matrix relating observations to specific combining ability (SCA) random effects $g_C \sim N(0, I\sigma_c^2)$ where g_C is a $c \times 1$ vector ($c = 3{,}792$), Z_A and Z_B are $n \times q_A$ and $n \times q_B$ design matrices relating observations to general combining ability (GCA) random effects for GA and GB, $g_A \sim N(0, A_A\sigma_A^2)$ and $g_B \sim N(0, A_B\sigma_B^2)$ respectively, where g_A and g_B are $q_A \times 1$ and $q_B \times 1$ vectors, respectively ($q_A = 2{,}037$ and $q_B = 340$). A_A and A_B are the pedigree-based kinship matrices of GA and GB, respectively.

In our work, we explored two types of distributions: binomial distribution, which is the appropriate one for proportional data, and normal distribution, for which more derived genetic parameters can be estimated.

The first model considers a binomial distribution such as:

$$Y_{c,t} \mid u_t, u_A, u_B, u_C \sim Bin\,(n_{c,t}, \pi_{c,t})$$

where $Y_{c,t}$ is the number of affected progenies in the cross (c) and the trial (t) among the number of inoculated progenies $n_{c,t}$, and $\pi_{c,t}$ is the associated probability.

The link function g is the logit such as:

$$g(\pi_{c,t}) = \log\left(\frac{\pi_{c,t}}{1-\pi_{c,t}}\right) = \eta_{c,t}$$

The second model considers a normal distribution such as:

$$Y_{c,t} \mid u_T, u_A, u_B, u_C \sim N\ (\eta_{c,t}, \sigma^2)$$

where $Y_{c,t}$ is the proportion of affected progenies in the cross (c) and the trial (t) , σ^2 is the residual variance, and the link function is the identity. Note that this second model is a linear mixed model (LMM).

Both models enabled prediction of the best linear unbiased predictor (BLUP) for each GB individual used in the QTL mapping, A_B being replaced by an identity matrix in order to avoid using the pedigree information that was subsequently used in the QTL analysis. Both statistical models were performed using ASReml-R software (Butler et al. 2007, V4) and resulted in two vectors of BLUP for group B individuals that were used in subsequent QTL mapping analysis.

2.3. Molecular data and genetic map construction

The 334 freeze-dried oil palm leaf samples available at the Cirad DNA-bank for the GB individuals included in the analysis were genotyped with 199 SSR markers developed in different studies. Among the 199 markers, 177 markers were developed by Cirad (Billotte et al. 2005), two by the Lee et al. (2015), four markers by the Malaysian Palm Oil Board (MPOB) (Zaki et al. 2012) and 18 expressed sequence tags markers were developed by IRD (*Institut de Recherche pour le Développement*) and Cirad (Tranbarger et al. 2012). These markers were selected based on a previous integrated pedigree-based genetic map constructed from a population of related individuals (Cochard et al. 2015). Selection was for a uniform distribution in the genome and the highest level of polymorphism in both LM and YBI genetic backgrounds. The information concerning markers was gathered in the supp. Table 2. DNA extraction, evaluation of the DNA concentrations and microsatellite fragment amplification were performed using the protocol described in Cochard et al. (2015). Genemapper© V4.1 (Applied Biosystems, USA) software was used to determine the size of the alleles.

Three genetic maps were constructed, one for each of LM and YBI population and one integrated map using the pedigree-based linkage mapping software CRI-MAP v2.4 (Green et al. 1990), as described in Cochard et al. (2015).

Consistency of marker calling across pedigrees and absence of spurious rates of double recombination events were checked using both CRI-MAP and FlexQTL™, and data were improved where necessary. Genetic maps were drawn using MapChart v2.0 software (Voorrips 2002) and are presented in Supporting Information Figure S1.

2.4. Pre-nursery QTL mapping approach

QTL mapping of *Ganoderma* disease resistance in pre-nursery conditions followed two main steps. The first step was carried out using a Bayesian approach and a multiple QTL model implemented in FlexQTL™ (Bink et al. 2002, 2014, 2008; www.flexqtl.nl) on the pre-nursery data after modeling, in order to identify putative QTL positions and predict the QTL genotypes. The second step consisted in stepwise QTL model selection on the raw pre-nursery data using the predicted QTL genotypes as fixed effects in the LMM.

2.4.1. QTL region identification and QTL genotype prediction

Six separate QTL analyses, corresponding to the two vectors of GB individual BLUP (see Phenotypic data section) with three different starting random seeds were performed using FlexQTL™. The six QTL analyses were based on a model with additive QTL effects, with the parameters MaximQTL and priorQTL set at 20 and 5 respectively for the Markov chain Monte Carlo simulation. The length of the Markov chains were set at 1 000 000 with a thinning value of 1 000. Using these parameters, the convergence indicators reached satisfying values for each parameter assessed (overall mean, μ, the residual variance, σ_e^2, the number of QTLs, $N_{QTL,}$ and the variance of QTLs, v_{QTL}). QTL regions were marked from the marginal posterior distributions of the six simulations and consensus QTL positions identified at the peaks of the summed posterior intensities profiles over the six simulations. QTL regions were named by the concatenation of population ID (LM, YBI or GB which refers to the grouped LM and YBI populations), the linkage group and the peaks separated by "@". For each consensus QTL, QTL genotypes for all individuals in the pedigree were predicted based on the vectors of QTL genotype posterior probabilities extracted from the FlexQTL output "MQTRegionsGTP.csv". QTL genotypes values were calculated as $[(0 * P_{qq}) + (1 * P_{Qq}) + (2 * P_{QQ})]$, with P the probability associated with the qq, qQ and QQ QTL genotypes, q being the favorable allele in this case. The continuous [0,2] values of the QTL genotypes were converted into discrete values {0,1,2} using the following threshold: values in the ranges [0,0.7],]0.7,1.3[and [1.3,2], were assigned to 0, 1 and 2 respectively, corresponding to individuals carrying homozygous favorable, heterozygous or homozygous unfavorable disease resistance alleles at the respective QTL regions considered.

2.4.2. Stepwise QTL model selection

In order to obtain a full QTL model fitted on the raw phenotypic data, QTL results from different modeling and random seeds were aggregated using stepwise model selection. The stepwise approach was applied on QTL genotypes vectors tested in the LMM model (see Phenotypic data section), following the procedure of the stepwiseqtl function of the R/qtl package (Broman and Sen 2009). First, a main effect QTL model was selected by testing the QTL genotype vectors in the LMM model with sequentially, a forward selection and a backward elimination. Model selection was based on the Akaike information criterion (AIC, Akaike 1998) using the full loglikelihood (Verbyla 2019). Similarly, the main effect QTL model was extended to the complete QTL model by first testing the interactions between QTLs and both QTL and CL (fixed effects), and second with the GA genetic background (random effect). Stepwise model selection was performed using ASReml-R software (Butler et al. 2007, V4).

2.5. Field evaluation of pre-nursery QTL

The relationships between *Ganoderma* genetic resistance in pre-nursery and field conditions were investigated using the census of disease status of the La Mé parents planted in genealogical gardens (see plant material section). The *Ganoderma* infection status was recorded biannually on 219 LM individuals planted in 1974 (5), 1976 (11), 1996 (5), 1997(107), 1998 (1), 1999 (47), 2001 (20) and 2003 (23) in six different blocks at Bangun Bandar estate, Indonesia. The disease status recording began within the three years after planting in the case of plantation after 1990 and in the 2000s for older plantings, and the last observation was recorded in 2018. *G. boninense* disease symptoms were scored blindly based on a six-level scale as described in Tisné et al. (2017). The appearance of the first *Ganoderma* symptom (T1S, first observation of score 2–6) was recorded and the associated time was considered as survival time, i.e., time from planting to the time the event occurred. The survival data were analyzed using the Cox model integrating a fixed effect for the date of planting:

$$\lambda(t, X) = \lambda_0(t) e^{X\beta} \quad (1)$$

where t is the time to the event or censoring, λ_0 denotes the baseline hazard function, X is the $n \times d$ design matrix relating the survival outcome for individuals to date of planting effects ($d = 8$) and $\beta = (\beta_1, \dots, \beta_d)$ is a $d \times 1$ unknown vector.

The effects of pre-nursery QTL were evaluated using the likelihood ratio test, for which the limiting distribution follows a chi-squared distribution, between the model (1) and the following model (2):

$$\lambda(t, X) = \lambda_0(t) e^{X\beta + X_q q} \quad (2)$$

with X_q being the {0,1,2} vector of pre-nursery-based QTL genotypes for the individuals and q the QTL effect. The analysis was performed with R software version 3.2.3 (Team 2012) and the *survival* package (Therneau 2015).

3. Results

3.1. Segregation of *Ganoderma* resistance in the GB population

Resistance to *Ganoderma* disease was tested in pre-nursery trials on 3,792 GA×GB crosses. On average, 30.8% of oil palm seedlings per cross presented disease symptoms at the end of the trial, ranging from 3 to 92.5% among the different crosses (Fig. 2a). Both LMM or GLMM models led to very similar predictions of GCA for the GB parents (r=0.97). Predictions of GCA were higher in YBI genetic background compared to LM, indicating higher susceptibility of the YBI background tested in this study (Fig 2b-c). Within genetic backgrounds, the distribution of GCA indicated segregation of quantitative resistance among founders, with mainly additive effects. Indeed, in LM genetic background, LM_1 self-pollinated individuals were the most resistant, and all the combinations of LM_1 and the alternative founders LM_2 or LM_3 showed higher resistance than the populations derived from self-pollinations of LM_2 and LM_3 (Fig. 2b-c). Similarly in YBI, YBI_3 was the least resistant genetic background, but its combination with YBI_2 improved the resistance of derived individuals. Even in narrow genetic bases, i.e. self-pollinated progenies of the most recent generation, there was still segregation of the resistance supporting the quantitative nature of *Ganoderma* resistance (Fig. 2b-c).

3.2. Genetic bases of *Ganoderma* resistance in pre-nursery trials

QTL mapping of the *Ganoderma* disease resistance in the GB population was performed using a Bayesian approach. Cumulating both modeling and the three random seeds per model, the number of QTLs was 125 considering all the marked QTL regions found by FlexQTL, regardless the 2lnBF threshold (supp. Table 3). These 125 QTL corresponded to around 20 QTLs on average per simulation. The QTLs were distributed in 30 consensus regions covering every linkage group (LG), with overall, a similar pattern between the different simulations (Fig. 3). Among these 29 QTL regions, 11 located on LG 1, 5, 6, 8, 9, 10, 12, 13 and 16 were identified consistently in the six simulations. The QTL mapping performed separately in LM and YBI revealed different QTL patterns between them: consistent QTL regions on LG 1, 6, 10, 12 and 13 segregated in the LM genetic background while the regions were located on LG 5, 8, 9 and 10 in the YBI genetic background (Supporting Information Figure S2). The average length of the QTL interval was around 25 cM (4-107 cM). Considering QTL genotypes in the 30

342 consensus QTL regions, there were on average, 35, 41 and 24% of *QQ*, *Qq* and *qq* genotypes respectively, in the
343 GB population, q being the favorable allele in this case.

344 Stepwise model selection was performed based on the QTL genotype vectors calculated for the 30 consensus QTL
345 regions. The first step fitted the LMM and indicated that the components related to the genetic effects represented
346 21% of total phenotypic variation, while 6% corresponded to the GCA of the GB individuals (Fig. 4). The final
347 QTL model retained four main effect QTL on LG 8, 9, 10 and 16, and one in interaction with the GA genetic
348 background on LG 6 (Fig. 4, supp. Table 4). Adding either the main effect or interacting QTLs in the LMM in the
349 different steps did not change the values of the non-genetic components, whereas the GCA_{GB} was reduced to 1%.
350 Including the interaction between the QTL on LG6 and the GA genetic background reduced both the values of the
351 SCA and the GCA_{GA} components. The partial determination coefficients computed for each QTL ranged from
352 0.05-2% of the total phenotypic variance, corresponding to 3-9% of genetic variance.

353 3.3. Effects of pre-nursery-based QTL on field *Ganoderma* resistance in the La Mé parents

354 The effects of the QTL identified using the pre-nursery data on GA×GB crosses were evaluated in the field where
355 219 LM parents included in the pre-nursery study were planted and underwent natural, uncontrolled *Ganoderma*
356 infection. The time of the first *Ganoderma* symptom appearance (T1S) was modeled using Cox regression with
357 the date of planting as covariate ($P < 0.01$). The effect of the percentage of favorable alleles per individual among
358 the 21 QTL regions identified in the LM genetic background (range 28-75%) was first assessed to evaluate the
359 global trend between pre-nursery and field conditions. The percentage of favorable alleles effect was not found to
360 be significant (P=0.2), but Kaplan-Meier estimates of survival showed consistency between the pre-nursery and
361 field QTL effects, a higher percentage of favorable alleles increased the probability of survival (Fig. 5a). Hence,
362 the individuals with less than 50% of favorable alleles were twice more affected by *Ganoderma* 20 years after
363 planting than individuals with more than 50% of favorable alleles (Fig. 5a). Then QTL genotype vectors, predicted
364 either GB or LM populations, were tested one at a time as covariates in the Cox model. The level of statistical
365 evidence of QTL effects between pre-nursery and field data was not correlated and significant QTL effects were
366 found for both a high (LG 9) or low (LG 4, 15) level of evidence in pre-nursery conditions (Fig. 5b). However the
367 direction of effects between field and pre-nursery effects was consistent for 78% of the QTLs, and for 89% when
368 a P-value=0.05 threshold was applied in the Cox model (Fig. 5b, Supporting Information Figure S3).

369

370 **4. Discussion**

Marker assisted selection (MAS) has a great potential for plant breeding and has been widely used for many crops with substantial achievements, especially for resistance to biotic stresses (Muranty et al. 2014). MAS should be particularly useful for perennial crops with a long breeding cycle and high phenotyping costs like oil palm, despite the identified biological, socioeconomic or technical issues (Muranty et al. 2014). In this paper, we report the proof of concept of an efficient *in silico* QTL mapping approach based on data collected in an ongoing breeding program. This allowed us to gain valuable insights into the genetic architecture of *Ganoderma* resistance and the transferability between field and pre-nursery results, as a basis for a future MAS.

4.1. Opportunities and issues of QTL mapping using data from breeding programs

Breeding programs for perennials are inherently geared towards long-term work with extensive data recording. This make them highly suited to the *in silico* approach, which is likely to improve the statistical properties of QTL detection through the increase in population size and diversity compared to conventional biparental populations. However, the specificity of the data from breeding programs, such as the extent of non-genetic effects due to long-term data or the genetic and phenotypic design unbalances due to the selection process, could reduce the expected benefits of QTL detection, namely its power and the accuracy of QTL location and QTL effect estimation (Würschum 2012). Hence, these datasets require a first stage of statistical modeling to account for several non-genetic effects and to obtain genotypic values. Thanks to their flexibility, mixed models are ideal tools to handle several types of data and effects (Smith et al. 2005). We used two types of mixed models, LMM and GLMM that enabled us to predict the GCA of genotyped individuals while accounting for confounding effects. We subsequently used these GCA values in FlexQTL because this software requires only one value per genotyped individual whereas they were progeny tested in the pre-nursery trials. Such a two-stage approach could affect QTL results so one-stage approaches are preferred when possible (Xue et al. 2017; Barrasso et al. 2019). The two types of mixed model used in this study did not lead to major differences in the QTLs identified, and a one-stage IBD–based variance component approach previously reported for production traits (IBD-VC, Tisné et al. 2015) that we used on pre-nursery *Ganoderma* data also produced similar results (data not shown). However, the calculation time requirement for the IBD-VC is an obstacle to a proper estimation of the significance threshold by permutation and a multi-QTL mapping procedure, which made us favor the approach presented.

Few studies have assessed the effects of the dataset features on QTL detection. In barley, using GWAS with an unbalanced dataset, the false positive rate was increased, whereas one-stage analysis performed better (Wang et al. 2012). In durum wheat, a GWAS performed both on an unbalanced and balanced dataset from a breeding

program showed major overlapping of selected SNP (Johnson et al. 2019). In diploid potato, a dataset grouping F3 families under selection was analyzed using either GWAS, stratified linkage or IBD based approaches that led to consistent QTL detection, but revealed issues concerning the QTL allele frequencies that could affect the results (Korontzis et al. 2020). In our study, the population studied could be genetically biased due to prior selection of the crosses tested for *Ganoderma* resistance based on yield related traits. However, inspection of QTL genotype frequencies showed that there were no depleted allelic classes among the QTL retained in the stepwise model selection. Moreover, the QTL genotype vectors predicted at the QTL regions were not correlated for the different linkage groups, indicating little segregation distortion that could have arisen due to the selection process.

Concerning the accuracy of QTL location, the increased population size allowed by the *in silico* approach should reduce the QTL interval thanks to the increased number of recombinations. In this proof of concept study, we chose to genotype the population using well characterized SSR markers in order to be able to connect the results with previous ones obtained with related populations. However, the QTL intervals were much larger than in other studies using FlexQTL on populations of similar size but with thousands of markers, indicating that the density was insufficient to mark them accurately. The large QTL regions could probably be considerably reduced thanks to the favorable genetic design and we are currently performing high-density SNP genotyping to achieve this objective. Beyond this limitation, the use of FlexQTL was particularly interesting: the use of IBD information mitigates the effect of low density genotyping, and the prediction of QTL genotypes offers the opportunity to use them in subsequent analyses. Hence, we were able to select a full QTL model using the raw data by testing main and interaction effects, and to assess the effects of pre-nursery QTL in the field. As reported by Verma and Whitaker (2018), QTL genotypes have great potential in the breeding context, for example, to predict QTL alleles for unobserved individuals in the breeding program based only on their marker and pedigree information, and then their expected resistance level.

4.2. Insights into the genetic architecture of *Ganoderma* resistance in oil palm

A first insight into genetic architecture came from the variance decomposition using the sire and dam mixed model designed for the analysis of the data on GA×GB hybrids. The genetic component, i.e. GCA in both heterotic groups and SCA, represented around 20% of the total phenotypic variance, which was expected due to the consistent genetic resistances identified in contrasted crosses or clones, balanced by the moderate repeatability of the screening tests (Durand-Gasselin et al. 2018). More surprising, the variance assigned to the GA pedigree was double that for the GB pedigree, while the pure parental GB genetic backgrounds are both more resistant and

exhibit more resistance variability than GA backgrounds (Durand-Gasselin et al. 2018). This could be an artefact of the unbalanced number of parents screened between heterotic groups and further investigation is needed to accurately estimate their relative contribution to the GA×GB resistance. The variance associated with SCA effect was 20% of the genetic variance and one QTL×genetic background interaction was retained, while well supported previous observations indicated that resistance was mainly additive, both in pre-nursery and field trials (Durand-Gasselin et al. 2018). Again, this could be an artefact, as only the GB pedigree was genotyped for this study but further analyses using both heterotic groups will allow us to estimate the proportion of variance due to GA×GB interaction and identifying underlying QTL.

The distributions of the GCA of GB individuals showed segregation of the *Ganoderma* resistance throughout the pedigree, even in the most inbred generations. Consequently, we identified a large number of putative QTL regions using FlexQTL, with weak to moderate effects. This partially reflects the composition of the GB that grouped two contrasted populations, LM and YBI, which displayed distinct QTL patterns when analyzed separately. However, even when we focused on a restricted genetic background, the large number of putative QTL found despite the reduced population size confirm the quantitative nature of *Ganoderma* resistance (quantitative disease resistance, QDR). Thus, the marked difference in *Ganoderma* resistance consistently observed between the four full-sib founders of the studied LM pedigree (Durand-Gasselin et al. 2018) is rather the consequence of a better combination of many favorable alleles than of a limited number of major QTLs. The numerous QTL found and the dissimilarity of QTL patterns between the LM and YBI genetic backgrounds is likely due to either the *Ganoderma* bio-trophic pathogenesis that induce contrasted transcriptomic responses (Bahari et al. 2018) or the multiple mechanisms involved in the QDR (Poland et al. 2009). This could explain the few discrepancies observed for some pre-nursery QTL with no effect in the field, and even a QTL with an opposite effect on LG12, considering that such QDR mechanisms are more prone to depend on the age of palms, on the environmental conditions, or on the genetic background surveyed.

Inspection of QTL colocalization may validate putative QTL when found for similar traits in independent experiments and inform QTL pleiotropy or linkage for different traits. Pleiotropy is especially worth investigating for QDR to obtain insights into possible underlying mechanisms and, together with linkage, on the resulting trade-off with other traits of interest (Nelson et al. 2018). To date, only two genetic mapping studies have been reported on *Ganoderma* resistance. The first analyzed data from a nursery test involving one resistant and two susceptible progenies, with a similar genetic background (Deli×YBI) and common markers to our study (Hama-Ali et al. 2015). Despite the limited scope of the study, i.e. involving only 79 individuals genotyped with 58 SSRs, Hama-

Ali et al. (2015) identified two significant markers on LG2 and seven in the same QTL regions as in our study, what is more, in equivalent populations, YBI and GB respectively. The second study used field data recorded on a multi-parental GA×GB population involving four GB founders that were the same as in the present study (*Eg*9PP population, Tisné *et al.* 2017). Four *Ganoderma* resistance loci were identified, two controlling the occurrence of the first *Ganoderma* symptoms (T1S), and two the death of palm trees (TD). Among them, the T1S QTL at the bottom of LG1 collocated with a QTL identified in GB and LM populations in the present study. The *Eg*9PP population and a large-scale genetic trial involving GB parents related to the founder of the present study (NGP population, Tisné *et al.* 2015, Tisné *et al.* 2019) were evaluated in the framework of the breeding program. Hence, data for fruit bunch production, oil extraction rate, and height increment traits were stored in databases, and both populations as well as the population from the present study were genotyped with the same SSR markers from a reference genetic map (Cochard et al. 2015) allowing QTL detection. We observed that among the six *Ganoderma* QTL regions with higher statistical support found in the GB, LM or YBI populations, most collocated with a large number of QTL for other agronomic traits (Tisné, personal communication). The colocalizations were more frequent in the LM population (33) than in the YBI one (15), while they were mostly found with oil extraction rate related traits and bunch number in LM genetic background in contrast with bunch weight and height increment in the YBI one (Tisné, personal communication). These preliminary findings now require further support, in particular by using a high-density SNP genotyping that is currently in progress, but already provide interesting insights into the possible diverse mechanisms underlying the QDR, which could differ considering the genetic backgrounds. This also highlights the benefits of the *in silico* approach assessed in this study that makes it possible to gather information from the entire breeding program for a more comprehensive description of the genetic architecture of traits of interest.

4.3. Advances towards a MAS of *Ganoderma* resistance in oil palm breeding programs

No complete resistance to *Ganoderma* has been identified to date and the results of the present study corroborate previous observations to indicate its quantitative nature (Franqueville et al. 2001; Idris et al. 2004; Durand-Gasselin et al. 2005). Despite the increasing use of QDR to improve the sustainability of disease resistance (Poland et al. 2009; Roux et al. 2014) the high number of loci and mechanisms involved makes its selection challenging. This is more acute in the case of oil palm with its long breeding cycle, worsened by the slow *Ganoderma* disease progression. Pre-nursery testing accelerated the screening of genetic material and revealed a genetic component that accounted for about 20% of phenotypic variance, which is generally a favorable level for a MAS perspective (Muranty et al. 2014). A first concern is to insure the consistency of QTL effects between the pre-nursery and field

489 results, like in conventional selection (Durand-Gasselin et al. 2018). We attempted to assess this at the QTL level
490 with the extensive use of the data from the breeding program, including the *Ganoderma* census routinely recorded
491 on seed and genealogical gardens. Following the previous study assessing the *Ganoderma* resistance in field we
492 used a survival analysis approach that provides several advantages (Tisné et al. 2017). Despite the limitations of
493 specific to the data recorded in seed gardens, i.e. mature palms of pure genetic backgrounds in the field vs GA×GB
494 seedlings in pre-nursey and spatio-temporal heterogeneity in the field, the accumulation of favorable pre-nursery
495 QTL alleles improved field resistance. Interestingly, the majority of QTL effect directions were consistent
496 regardless the statistical evidence in pre-nursery. Thus, the many QTL that would not have been detected in the
497 field setup because of a lack of statistical power, were identified in the pre-nursery study and are valuable for a
498 marker-assisted *Ganoderma* resistance selection.

499 Secondly, the quantitative nature of *Ganoderma* resistance identified could hamper the conventional QTL
500 pyramiding approach due to the high number of loci involved, especially considering the long generation time in
501 oil palm. In such a QDR context, the MAS approaches developed for other agronomic quantitative traits are
502 probably more suitable, especially the genomic selection (GS) approach (Poland and Rutkoski 2016). In oil palm,
503 GS has emerged as an efficient MAS method and is being increasingly evaluated for yield improvement (Nyouma
504 et al. 2019). Thus GS statistical models and implementation modes already assessed in oil palm could be
505 transferred or adapted to *Ganoderma* disease related data from the breeding program (Cros et al. 2015, 2017).
506 However, the qualitative/quantitative nature of disease resistance is a continuum (Poland et al. 2009). Despite a
507 large number of QTL regions identified using FlexQTL, only 5 QTL with weak to moderate effects explained
508 almost all the GB GCA component based on pre-nursery data. GS models including information on QTL or genes
509 have been proposed to improve prediction capacity in such situations (Bernardo 2014; Zhang et al. 2014) and
510 should be considered for a GS of implementation in light of the emerging insights into the genetic architecture of
511 *Ganoderma* resistance.

512 A final issue is that selection for *Ganoderma* resistance will need to be combined with resistance to other diseases
513 and cannot be at the expense of other traits of interests. The cost of disease resistance through negative trade-off
514 with performance or fitness was a long-lasting question in model plants but was less investigated in plant breeding
515 (Brown 2002). In the former section, we described colocalization of *Ganoderma* resistance QTL with yield related
516 ones, with a genetic background specificity of these complex patterns. Dealing with multiple traits and multiple
517 genetic background is challenging and the QTL information provided by the *in silico* approach assessed in the
518 present study is very valuable for comprehensive modeling of a MAS strategy. Hence, a recent study in oil palm

519 simulated the outcomes of alternative selection strategies on yield and its components based on their global genetic
520 architecture, including the pleiotropy/linkage and phases between the underlying QTL (Tisné et al. 2019). Virtual
521 individuals and crosses were simulated from the actual founders via meiosis simulations based on the QTL
522 positions identified with FlexQTL, which thus integrated their recombination frequencies. The QTL genotypes
523 predicted in FlexQTL enabled prediction of their multiple trait values and their incorporation in yield based on the
524 QTL effects. This use of QTL genotypes is of prime interest as QTL genotypes can be predicted based on markers
525 alone in any related individual, whether phenotyped or not. In the MAS perspective for *Ganoderma* resistance,
526 this approach would help attenuate possible trade-offs with other traits of interest and optimize the combination of
527 QDR from diverse genetic backgrounds.

528 **5. Conclusion**

529 The cost-effective and efficient *in silico* mapping approach assessed in this study has great potential for the
530 implementation of MAS of traits of interest in oil palm. Its application in the context of *Ganoderma* disease
531 resistance enabled us to use the considerable quantities of data generated in the framework of conventional
532 phenotypic selection to obtain valuable information in the MAS perspective. First, important information on the
533 genetic architecture of resistance to *Ganoderma* disease was obtained, confirming its quantitative nature and
534 identifying the loci involved. In addition, together with other ongoing works, this study sheds light on the
535 relationships between *Ganoderma* resistance and yield related traits that could produce undesirable trade-offs.
536 Second, the consistency between genetic resistance in pre-nursery conditions and in the field was assessed at the
537 QTL level and globally indicated satisfactory portability. However, a few loci deserve careful consideration due
538 to underlying mechanisms that could lead to contrasted phenotypic expression between pre-nursery and field
539 conditions. Finally, this proof-of-concept study provides guidelines for future works on *Ganoderma* disease
540 resistance and should encourage oil palm breeders to use this approach to collectively acquire a better
541 comprehension of its complex genetic architecture.

542 **Declaration of Competing Interest**

543 The authors declare that they have no conflict of interests.

544 **Data availability**

545 The datasets generated and analyzed during the current study are available from the corresponding author.

546 **Acknowledgments**

This study was based on a very intensive and laborious work involving many people in the long-term. We thank Zulkifi Lubis, Augustiaman Purba, Shri Jeweyen, and all the SOCFIN Indonesia staff at Tanah Gambus who performed the pre-nursery trials. We thank the PalmElit staff, Hubert de Franqueville and Michaël Pernaci for information on plant pathology and Nicolas Turnbull on breeding. We acknowledge Tristan Durand-Gasselin (PalmElit) for his insightful review of the study and manuscript. We thank Eric van de Weg (Wageningen UR) for the review of the manuscript.

This research was partly funded by a grant from PalmElit SAS. MD contributed partly to this study while she was visiting researcher at Georgetown University and supported by the European Union's Horizon 2020 research and innovation program under grant agreement No840383.

Supplementary Information

Supporting Information Figure S1: Genetic map of the prenursery GB, LM and YBI oil palm populations.

Supporting Information Figure S2: QTL mapping of the *Ganoderma* resistance in the prenursery LM and YBI oil palm populations.

Supporting Information Figure S3: Survival curves of the La Mé population in field conditions according to the genotypes of QTL identified based on the pre-nursery data.

References

Al-Obaidi JR, Mohd-Yusuf Y, Razali N, et al (2014) Identification of proteins of altered abundance in oil palm infected with Ganoderma boninense. Int J Mol Sci 15:5175–5192

Anciro A, Mangandi J, Verma S, et al (2018) FaRCg1: a quantitative trait locus conferring resistance to Colletotrichum crown rot caused by Colletotrichum gloeosporioides in octoploid strawberry. Theor Appl Genet 131:2167–2177. https://doi.org/10.1007/s00122-018-3145-z

Bahari MNA, Sakeh NM, Abdullah SNA, et al (2018) Transciptome profiling at early infection of Elaeis guineensis by Ganoderma boninense provides novel insights on fungal transition from biotrophic to necrotrophic phase. BMC Plant Biol 18:1-25. https://doi.org/10.1186/s12870-018-1594-9

Barrasso C, Memah M-M, Génard M, Quilot-Turion B (2019) Model-based QTL detection is sensitive to slight modifications in model formulation. PLOS ONE 14:e0222764. https://doi.org/10.1371/journal.pone.0222764

579 Bernardo R (2014) Genomewide selection when major genes are known. Crop Sci 54:68–75

580 Billotte N, Marseillac N, Risterucci A-M, et al (2005) Microsatellite-based high density
581 linkage map in oil palm (Elaeis guineensis Jacq.). TAG Theor Appl Genet Theor
582 Angew Genet 110:754–765. https://doi.org/10.1007/s00122-004-1901-8

583 Bink M, Uimari P, Sillanpää J, et al (2002) Multiple QTL mapping in related plant
584 populations via a pedigree-analysis approach. TAG Theor Appl Genet Theor Angew
585 Genet 104:751–762. https://doi.org/10.1007/s00122-001-0796-x

586 Bink MCAM, Anderson AD, van de Weg WE, Thompson EA (2008) Comparison of marker-
587 based pairwise relatedness estimators on a pedigreed plant population. Theor Appl
588 Genet 117:843–855. https://doi.org/10.1007/s00122-008-0824-1

589 Bink MCAM, Jansen J, Madduri M, et al (2014) Bayesian QTL analyses using pedigreed
590 families of an outcrossing species, with application to fruit firmness in apple. Theor
591 Appl Genet 127:1073–1090. https://doi.org/10.1007/s00122-014-2281-3

592 Breton F, Hasan Y, Hariadi, et al (2006a) Characterization of parameters for the development
593 of an early screening test for basal stem rot tolerance in oil palm progenies. In:
594 Technol. Breakthr. Commer. Way Forw. Proc. PIPOC 2005 Int. Palm Oil Congr.
595 Agric. Biotechnol. Sustain. 25-29 Sept. 2005 Petaling Jaya Malays.
596 http://agritrop.cirad.fr/543369/. Accessed 31 Jan 2018

597 Breton F, Hasan Y, Hariadi S, et al (2006b) Characterization of parameters for the
598 development of an early screening test for basal stem rot tolerance in oil palm
599 progenies. J Oil Palm Res 2006:24–36

600 Breton F, Rahmaningsih MR, Lubis Z, et al (2009) Early Screening Test: A Routine Work to
601 Evaluate Resistance/Susceptibility Level of Oil Palm Progenies to Basal Stem Rot
602 Disease. In. MPOB International Palm Oil Congress (PIPOC 2009), Kuala Lumpur, 9-
603 12 November 2009. MPOB.

604 Butt TM, Wang C, Shah FA, Hall R (2006) Degeneration of entomogenous fungi. In: An
605 ecological and societal approach to biological control. Springer, pp 213–226

606

607 Cochard B, Carrasco-Lacombe C, Pomiès V, et al (2015) Pedigree-based linkage map in two
608 genetic groups of oil palm. Tree Genet Genomes 11:1–12.
609 https://doi.org/10.1007/s11295-015-0893-7

610 Corley RHV (2009) How much palm oil do we need? Environ Sci Policy 12:134–139.
611 https://doi.org/10.1016/j.envsci.2008.10.011

612 Cros D, Bocs S, Riou V, et al (2017) Genomic preselection with genotyping-by-sequencing
613 increases performance of commercial oil palm hybrid crosses. BMC Genomics
614 18(1):1-17. https://doi.org/10.1186/s12864-017-4179-3

615 Cros D, Denis M, Sánchez L, et al (2015) Genomic selection prediction accuracy in a
616 perennial crop: case study of oil palm (Elaeis guineensis Jacq.). Theor Appl Genet
617 128:397–410

De Franqueville H, Renard JL (1990) Improvement of oil palm vascular wilt tolerance-results and development of the disease at the R. Michaux plantation. Oleagineux Vol.45 No.10 pp.399-405.

Durand-Gasselin T, Asmady H, Flori A, et al (2005) Possible sources of genetic resistance in oil palm (Elaeis guineensis Jacq.) to basal stem rot caused by Ganoderma boninense– prospects for future breeding. Mycopathologia 159:93–100

Durand-Gasselin T, Cochard B, de Franqueville H (2018) Advances in disease-resistant oil palm varieties. In: Center for International Cooperation in Agricultural Research for Development (CIRAD), France, Rival A (eds) Burleigh Dodds Series in Agricultural Science. Burleigh Dodds Science Publishing, pp 137–164

Faizah R, Putranto RA, Wening S, et al (2020) Differential expression of root specific genes of oil palm seedlings at early stage of *Ganoderma boninense* infection. IOP Conf Ser Earth Environ Sci 418:012044. https://doi.org/10.1088/1755-1315/418/1/012044

Franqueville H de, Asmady H, Jacquemard JC, et al (2001) Indications on sources of oil palm (Elaeis guineensis Jacq.) genetic resistance and susceptibility to Ganoderma sp., the cause of basal stem rot. In: Cutting-edge technologies for sustained competitiveness: Proceedings of the 2001 PIPOC International Palm Oil Congress, Agriculture Conference, Kuala Lumpur, Malaysia, 20-22 August 2001. Malaysian Palm Oil Board (MPOB), pp 420–431

Garin V, Wimmer V, Mezmouk S, et al (2017) How do the type of QTL effect and the form of the residual term influence QTL detection in multi-parent populations? A case study in the maize EU-NAM population. Theor Appl Genet 130:1753–1764. https://doi.org/10.1007/s00122-017-2923-3

Gascon JP, De Berchoux CH (1964) Caractéristiques de la production d'Elaeis guineensis (Jacq.) de diverses origines et leurs croisements. Appl À Sélection Palmier À Huile Oléagineux 19:75–84

George AW, Visscher PM, Haley CS (2000) Mapping quantitative trait loci in complex pedigrees: a two-step variance component approach. Genetics 156:2081–2092

Govender N, Abu-Seman I, Mui-Yun W (2020) Root Lignin Composition and Content in Oil Palm (Elaeis guineensis Jacq.) Genotypes with Different Defense Responses to Ganoderma boninense. Agronomy 10:1487. https://doi.org/10.3390/agronomy10101487

Govender NT, Mahmood M, Seman IA, Wong M-Y (2017) The Phenylpropanoid Pathway and Lignin in Defense against Ganoderma boninense Colonized Root Tissues in Oil Palm (Elaeis guineensis Jacq.). Front Plant Sci 8:1395

Green P, Falls K, Crooks S (1990) CRIMAP Documentation. https://www.animalgenome.org/hu/CRIMAPwkshp/crimap-doc.html. Accessed 5 Apr 2018

Hama-Ali EO, Panandam JM, Tan SG, et al (2015) Association between basal stem rot disease and simple sequence repeat markers in oil palm, Elaeis guineensis Jacq. Euphytica 202:199–206

659 Ho C-L, Tan Y-C, Yeoh K-A, et al (2016) De novo transcriptome analyses of host-fungal
660 interactions in oil palm (Elaeis guineensis Jacq.). BMC Genomics 17(1):1-19.

661 Idris A, Kushairi A, Ismail S, Ariffin D (2004) Selection for partial resistance in oil palm
662 progenies to Ganoderma basal stem rot. J Oil Palm Res 16:12–18

663 Jeennor S, Volkaert H (2014) Mapping of quantitative trait loci (QTLs) for oil yield using
664 SSRs and gene-based markers in African oil palm (Elaeis guineensis Jacq.). Tree
665 Genet Genomes 10:1–14

666 Johnson M, Kumar A, Oladzad-Abbasabadi A, et al (2019) Association Mapping for 24 Traits
667 Related to Protein Content, Gluten Strength, Color, Cooking, and Milling Quality
668 Using Balanced and Unbalanced Data in Durum Wheat [Triticum turgidum L. var.
669 durum (Desf).]. Front Genet 10:. https://doi.org/10.3389/fgene.2019.00717

670 Korontzis G, Malosetti M, Zheng C, et al (2020) QTL detection in a pedigreed breeding
671 population of diploid potato. Euphytica 216(9):1-14. https://doi.org/10.1007/s10681-
672 020-02674-y

673 Lee M, Xia JH, Zou Z, et al (2015) A consensus linkage map of oil palm and a major QTL for
674 stem height. Sci Rep 5:(1), 1-7.

675 Mangandi J, Verma S, Osorio L, et al (2017) Pedigree-based analysis in a multiparental
676 population of octoploid strawberry reveals QTL alleles conferring resistance to
677 Phytophthora cactorum. G3 Genes Genomes Genet 7:1707–1719

678 Mercière M, Laybats A, Carasco-Lacombe C, et al (2015) Identification and development of
679 new polymorphic microsatellite markers using genome assembly for Ganoderma
680 boninense, causal agent of oil palm basal stem rot disease. Mycol Prog 14:103

681 Meunier J, Gascon JP (1972) Le schéma général d'amélioration du palmier à huile à l'IRHO.
682 Oléagineux 27:1–12

683 Muranty H, Jorge V, Bastien C, et al (2014) Potential for marker-assisted selection for forest
684 tree breeding: lessons from 20 years of MAS in crops. Tree Genet Genomes 10:1491–
685 1510

686 Nelson R, Wiesner-Hanks T, Wisser R, Balint-Kurti P (2018) Navigating complexity to breed
687 disease-resistant crops. Nat Rev Genet 19:21–33. https://doi.org/10.1038/nrg.2017.82

688 Nusaibah SA, Akmar ASN, Idris AS, et al (2016) Involvement of metabolites in early defense
689 mechanism of oil palm (Elaeis guineensis Jacq.) against Ganoderma disease. Plant
690 Physiol Biochem 109:156–165

691 Nyouma A, Bell JM, Jacob F, Cros D (2019) From mass selection to genomic selection: one
692 century of breeding for quantitative yield components of oil palm (Elaeis guineensis
693 Jacq.). Tree Genet Genomes 15(5):1-16. https://doi.org/10.1007/s11295-019-1373-2

694 Paterson RRM (2019) Ganoderma boninense Disease of Oil Palm to Significantly Reduce
695 Production After 2050 in Sumatra if Projected Climate Change Occurs.
696 Microorganisms 7:24

Paterson RRM (2007) Ganoderma disease of oil palm—A white rot perspective necessary for integrated control. Crop Prot 26:1369–1376. https://doi.org/10.1016/j.cropro.2006.11.009

Poland J, Rutkoski J (2016) Advances and Challenges in Genomic Selection for Disease Resistance. Annu Rev Phytopathol 54:79–98. https://doi.org/10.1146/annurev-phyto-080615-100056

Poland JA, Balint-Kurti PJ, Wisser RJ, et al (2009) Shades of gray: the world of quantitative disease resistance. Trends Plant Sci 14:21–29

Pootakham W, Jomchai N, Ruang-areerate P, et al (2015) Genome-wide SNP discovery and identification of QTL associated with agronomic traits in oil palm using genotyping-by-sequencing (GBS). Genomics 105:288–295

Rees RW, Flood J, Hasan Y, et al (2009) Basal stem rot of oil palm (Elaeis guineensis); mode of root infection and lower stem invasion by Ganoderma boninense. Plant Pathol 58:982–989

Rees RW, Flood J, Hasan Y, Cooper RM (2007) Effects of inoculum potential, shading and soil temperature on root infection of oil palm seedlings by the basal stem rot pathogen Ganoderma boninense. Plant Pathol 56:862–870. https://doi.org/10.1111/j.1365-3059.2007.01621.x

Rochmyaningsih D (2019) Making peace with oil palm. Science 365:112–115. https://doi.org/10.1126/science.365.6449.112

Roux F, Voisin D, Badet T, et al (2014) Resistance to phytopathogens *e tutti quanti* : placing plant quantitative disease resistance on the map: Quantitative disease resistance in plants. Mol Plant Pathol 15:427–432. https://doi.org/10.1111/mpp.12138

Sakeh NM, Abdullah SNA, Bahari MNA, et al (2020) EgJUB1 and EgERF113 transcription factors as master regulators of defense response in Elaeis guineensis against the hemibiotrophic *Ganoderma boninense*. BMC plant biology, 21(1), 1-20.

Singh R, Low E-TL, Ooi LC-L, et al (2013) The oil palm SHELL gene controls oil yield and encodes a homologue of SEEDSTICK. Nature 500:340–344. https://doi.org/10.1038/nature12356

Smith AB, Cullis BR, Thompson R (2005) The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. J Agric Sci 143:449–462. https://doi.org/10.1017/S0021859605005587

Tan Y-C, Yeoh K-A, Wong M-Y, Ho C-L (2013) Expression profiles of putative defence-related proteins in oil palm (Elaeis guineensis) colonized by Ganoderma boninense. J Plant Physiol 170:1455–1460. https://doi.org/10.1016/j.jplph.2013.05.009

Team RC (2012) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2012. ISBN 3-900051-07-0

Tee S-S, Tan Y-C, Abdullah F, et al (2013) Transcriptome of oil palm (Elaeis guineensis Jacq.) roots treated with Ganoderma boninense. Tree Genet Genomes 9:377–386

736 Tisné S, Denis M, Cros D, et al (2015) Mixed model approach for IBD-based QTL mapping
737 in a complex oil palm pedigree. BMC Genomics 16(1):1-12.

738 Tisné S, Maurin G, Bink M, et al (2019) Complex Trait Improvement in the Reciprocal
739 Recurrent Selection Context using a Pedigree Based QTL Mapping Approach. In:
740 Proceedings of the PIPOC 2019 International Palm Oil Congress Agriculture,
741 Biotechnology & Sustainability Conference. Malaysian Palm Oil Board (MPOB),
742 Kuala Lumpur Convention Centre, Kuala Lumpur, Malaysia, pp 356–362

743 Tisné S, Pomiès V, Riou V, et al (2017) Identification of Ganoderma disease resistance loci
744 using natural field infection of an oil palm multiparental population. G3 Genes
745 Genomes Genet 7:1683–1692

746 Tranbarger TJ, Kluabmongkol W, Sangsrakru D, et al (2012) SSR markers in transcripts of
747 genes linked to post-transcriptional and transcriptional regulatory functions during
748 vegetative and reproductive development of Elaeis guineensis. BMC Plant Biol
749 12(1):1-12. https://doi.org/10.1186/1471-2229-12-1

750 van de Weg E, Di Guardo M, Jänsch M, et al (2018) Epistatic fire blight resistance QTL
751 alleles in the apple cultivar 'Enterprise' and selection X-6398 discovered and
752 characterized through pedigree-informed analysis. Mol Breed 38(1):1-18.
753 https://doi.org/10.1007/s11032-017-0755-0

754 van de Weg WE, Voorrips RE, Finkers R, et al (2004) Pedigree genotyping: a new pedigree-
755 based approach of QTL identification and allele mining. In Acta Hortic 45–50.
756 https://doi.org/10.17660/ActaHortic.2004.663.1

757 van Eeuwijk FA, Boer M, Totir LR, et al (2010) Mixed model approaches for the
758 identification of QTLs within a maize hybrid breeding program. Theor Appl Genet
759 120:429–440

760 Verma S, Whitaker VM (2018) Prediction of QTL genotypes and trait phenotypes using
761 FlexQTLTM: a pedigree-based analysis approach. J. Plant Biol. Crop Res, 2, 1006.

762 Voorrips RE (2002) MapChart: Software for the Graphical Presentation of Linkage Maps and
763 QTLs. J Hered 93:77–78. https://doi.org/10.1093/jhered/93.1.77

764 Woittiez LS, van Wijk MT, Slingerland M, et al (2017) Yield gaps in oil palm: A quantitative
765 review of contributing factors. Eur J Agron 83:57–77

766 Würschum T (2012) Mapping QTL for agronomic traits in breeding populations. Theor Appl
767 Genet 125:201–210

768 Xue S, Ogut F, Miller Z, et al (2017) Comparison of one-stage and two-stage genome-wide
769 association studies. bioRxiv 099291. https://doi.org/10.1101/099291

770 Yeoh K-A, Othman A, Meon S, et al (2013) Sequence analysis and gene expression of
771 putative oil palm chitinase and chitinase-like proteins in response to colonization of
772 Ganoderma boninense and Trichoderma harzianum. Mol Biol Rep 40:147–158.
773 https://doi.org/10.1007/s11033-012-2043-8

774 Yeoh K-A, Othman A, Meon S, et al (2012) Sequence analysis and gene expression of
775 putative exo- and endo-glucanases from oil palm (Elaeis guineensis) during fungal
776 infection. J Plant Physiol 169:1565–1570. https://doi.org/10.1016/j.jplph.2012.07.006

777 Zaki NM, Singh R, Rosli R, Ismail I (2012) Elaeis oleifera Genomic-SSR Markers:
778 Exploitation in Oil Palm Germplasm Diversity and Cross-Amplification in Arecaceae.
779 Int J Mol Sci 13:4069–4088. https://doi.org/10.3390/ijms13044069

780 Zhang Z, Ober U, Erbe M, et al (2014) Improving the accuracy of whole genome prediction
781 for complex traits using the results of genome wide association studies. PloS One
782 9:e93017

783

784 **Figure captions**

785 **Fig. 1** Pedigree of the pre-nursery GB oil palm population. Boxes on the left represent the founders of the La Mé
786 (LM, panel A) and Yangambi (YBI, panel B) populations. Note that the La Mé founders LM_1:4 are full sibs.
787 Other boxes represent full-sib families whose color represents their relation to their genetic background, with the
788 number of individuals in parenthesis. The circled cross symbols represent progenies obtained through self-
789 pollination, and successive self-pollinated progenies keep the same color.

790 **Fig. 2** Distribution of *Ganoderma* disease resistance in the pre-nursery GB oil palm population. Distribution of
791 the percentage of affected individuals in crosses (A), BLUP obtained from random effect of the GCA in GB in a
792 GLMM (B) and LMM (C) for the La Mé (LM) and Yangambi (YBI) populations. Different colors represent
793 different genetic backgrounds.

794 **Fig. 3** QTL mapping of *Ganoderma* resistance in the pre-nursery GB oil palm population. QTL regions marked
795 by FlexQTL software in six independent simulations (LMM and GLMM models, three random starting seeds) (A)
796 and the averaged posterior intensity calculated at a 1 cM grid for the six simulations (B) are plotted along the
797 genome. In panel A, the yellow to red color code scale depict the value of intensity of the corresponding marked
798 QTL regions found in the "MQTRegions.new" FlexQTL output file. In panel B, a white to red color scale indicates
799 the number of marked QTL regions among the six simulations at the corresponding position in the genome.

800 **Fig. 4** Variance components of *Ganoderma* resistance in the pre-nursery screening tests. Variance components are
801 plotted as a percentage of the total phenotypic variance for each of the steps performed in the stepwise selection
802 model. GA/GB: heterotic group A and B; GCA: general combining ability; SCA: Specific combining ability; CL:
803 *Ganoderma* clonal lines; QTL names: see M&M section.

Fig. 5 Pre-nursery QTL effects on *Ganoderma* resistance to natural field infection in the La Mé genetic background. (A) Survival curves of the La Mé population according to the percentage of favorable alleles at the 21 La Mé QTL detected in the pre-nursery analysis, the red to green color scale indicates an increasing percentage. Survival estimates are plotted at the time of the first observation of a *Ganoderma* symptom. (B) Scatterplot showing the relationship between the statistical significances of QTL effects in the pre-nursery experiments (posterior intensity, x-axis) and in the field (-log (P-value) from the Cox model, y-axis). QTL originate from QTL mapping using the GB (squares) or LM (triangles) pedigree. Consistency between field and pre-nursery QTL effects was defined for QTL alleles decreasing the number of affected progenies in the pre-nursery trials and delaying the appearance of the first symptom of *Ganoderma*: inconsistent and consistent QTL effects are depicted by green (+) or red (-) symbols, respectively. QTL for which one of the three allelic classes (*QQ*, *Qq* or *qq*) was represented by less than ten individuals are depicted by shaded symbols. QTL names: see M&M section.

Supplementary figure S2

Supplementary figure S1

Supplementary figure S1

Supplementary table 1

Supplementary table 2

Click here to access/download

Supplementary Material

Supplementary_table2.csv

Supplementary table 3

Click here to access/download

Supplementary Material

Supplementary_table3.csv

Supplementary table 4

Click here to access/download

Supplementary Material

Supplementary_table4.csv

R script

Click here to access/download

Supplementary Material

Rscript_Surv.R