The ash dieback invasion of Europe was founded by two genetically divergent individuals

Mark McMullan[*,1], **Maryam Rafiqi**[2], **Gemy Kaithakottil**[1], **Bernardo Clavijo**[1], **Lorelei Bilham**[3], **Elizabeth Orton**[3], **Lawrence Percival-Alwyn**[1], **Ben J. Ward**[1], **Anne Edwards**[3], **Diane G.O. Saunders**[3], **Gonzalo Garcia**[1], **Jonathan Wright**[1], **Walter Verweij**[1], **Georgios Koutsovoulos**[4], **Kentaro Yoshida**[5,6], **Tsuyoshi Hosoya**[7], **Louisa Williamson**[8], **Philip Jennings**[8], **Renaud Ioos**[9], **Claude Husson**[10], **Ari M. Hietala**[11], **Adam Vivian-Smith**[11], **Halvor Solheim**[11], **Dan MaClean**[5], **Christine Fosker**[1], **Neil Hall**[1], **James K.M. Brown**[3], **David Swarbreck**[1], **Mark Blaxter**[4,12], **Allan Downie**[3], and **Matthew D. Clark**[*,1,13]

[1]The Earlham Institute, Norwich Research Park, Norwich, NR4 7UZ, UK

[2]Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3AE, UK

[3]John Innes Centre, Norwich Research Park, Norwich, NR4 7UH, UK

[4]Institute of Evolutionary Biology, The University of Edinburgh, Edinburgh, EH9 3FL, UK

[5]The Sainsbury Laboratory, Norwich Research Park, Norwich, NR4 7UH, UK

[6]Graduate school of Agricultural Science, Kobe University, Kobe, Hyogo, Japan

[7]Department of Botany, National Museum of Nature and Science, Tsukuba, Ibaraki, 305-0005 Japan

[8]Fera Science Limited, Sand Hutton, York, YO41 1LZ, UK

[9]ANSES Laboratoire de la Santé des Végétaux, 54220 Malzéville, France

[10]UMR IAM, INRA, Université de Lorraine, 54000 Nancy, France

[11]Norwegian Institute of Bioeconomy Research, Høgskoleveien 8, 1433 Ås, Norway

[12]Edinburgh Genomics, School of Biological Sciences, University of Edinburgh, Edinburgh, EH9 3JT, UK

[13]Natural History Museum, Department of Life Sciences, London SW7 5BD, UK

Abstract

Accelerating international trade and climate change make pathogen spread an increasing concern. *Hymenoscyphus fraxineus*, the causal agent of ash dieback, is a fungal pathogen that has been moving across continents and hosts from Asian to European ash. Most European common ash trees (*Fraxinus excelsior*) are highly susceptible to *H. fraxineus*, although a minority (~5%) have partial resistance to dieback. Here, we assemble and annotate a *H. fraxineus* draft genome which approaches chromosome scale. Pathogen genetic diversity across Europe and in Japan, reveals a strong bottleneck in Europe, though a signal of adaptive diversity remains in key host interaction genes. We find that the European population was founded by two divergent haploid individuals. Divergence between these haplotypes represents the ancestral polymorphism within a large source population. Subsequent introduction from this source would greatly increase adaptive potential of the pathogen. Thus, further introgression of *H. fraxineus* into Europe represents a potential threat and Europe-wide biological security measures are needed to manage this disease.

Human movement and climate change may increase the incidence of plant pathogen introductions as new environments become favourable[1]. Dutch elm disease is one example of a pathogen epidemic which emerged and caused the loss of billions of elm trees globally[2]. Presently, the ascomycete fungus, *Hymenoscyphus fraxineus*, is causing disease and death in European common ash, *Fraxinus excelsior* and *F. angustifolia* (narrow-leaved ash). *H. fraxineus* jumped host from Asian ash species where it is a leaf pathogen with little impact on its host. In Europe it is killing ash at an alarming rate and displacing the non-aggressive indigenous fungus, *H. albidus*[3].

The disease was first observed in Europe in north-western Poland in 1992 and, moving west, was identified in the UK in 2012[3]. Less than 5% of trees are partially resistant or tolerant[4,5] to ash dieback disease which is characterised by dark brown/orange leave lesions on leaves followed by wilting, necrotic lesions on shoots, then diamond-shaped lesions on the stems and finally dieback of the crown[6] (Fig. 1). The loss of leaves in the crown of mature trees proceeds over years and leads, in severe cases, to tree death. *H. fraxineus* is a heterothallic fungus[7] which has been shown to reproduce asexually *in vitro*[8] but in the wild, sexual reproduction occurs annually on fallen leaf rachises in the leaf litter[6] (Fig. 1).

A pathogen's evolutionary potential, is rooted in its genetic diversity (or its effective population size)[9]. All else being equal, large populations can adapt more quickly than small ones for two reasons; first, a large population carries more polymorphism and so there is a greater chance that a favourable mutation is segregating. Second, the impact of random genetic drift is lower for larger populations and therefore, natural selection is better able to drive those favourable mutations to fixation[10]. However, pathogen introductions are often associated with genetic bottlenecks, or founder effects which reduce the level of genetic diversity and the efficacy of selection. This disparity between reduced polymorphism and invasion success is known as the genetic paradox of biological invasions[11]. The potato late blight pathogen, *Phytophthora infestans*, is another example of a successful bottlenecked pathogen introduction[12]. However devastating a pathogen may be, multiple introductions

can increase diversity higher than that of native populations[11]. Early *P. infestans* invasions were dominated by a single clonal lineage (US-1) which was later superseded as diversity and severity increased[12]. The Dutch elm disease pandemic(s) also had an initial rapid spread of *Ophiostoma ulmi* which was later replaced by O. novo-ulmi (also subsp. *americana*), and went on to devastate elm populations across two continents[2]. The amount of *H. fraxineus* genetic variation introduced to Europe and, the potential for further introduction, is therefore of critical concern.

Estimates of *H. fraxineus* microsatellite allelic richness suggest as few as two haploid individuals may have invaded Europe[13]. Interestingly, the viral pathogen of *H. fraxineus* (mitovirus 1) also has two genetic groups present in the European population[14]. Given the impact of the pathogen invasion so far, a founder effect of just two haploid individuals would represent a genetic paradox and so it is important to measure genetic diversity across the genome as well as meaningful, adaptive diversity in the host interaction genes (effectors). Determining the genetic diversity present in the native range is key to understanding the consequences of any further introductions. Here, we assembled and annotated a high-quality draft genome of *H. fraxineus*. We quantified the level of genetic variation in 43 *H. fraxineus* haploid isolates from across Europe as well as 15 haplotypes from part of its native range (Japan). To understand the adaptive potential of the ash dieback pathogen in Europe as well as the potential of future invasions, we determined the effective population size, the bottleneck into Europe and estimated the size of the source population. In addition, we mined putative effector genes from the genome of *H. fraxineus*. These genes encode secreted proteins considered to play key role in establishing fungal infections and facilitating disease development[15]. We measured signals of adaptive diversity in the putative effectors and all other genes. The proportion of introduced adaptive variation and the potential for further introduction are a key focus of the present work.

Results & Discussion

Population genetic diversity of the emerging invasive *H. fraxineus* pathogen measured among 43 isolates from across Europe is an eighth of that observed in 15 haplotypes from a single wood in Japan. This general reduction in polymorphism, caused by a bottleneck of two divergent individuals, could reflect a reduction in the pathogens adaptive potential in its introduced environment. This reduction in genetic diversity is present in effectors as well as all other genes. European effectors retain a greater degree of adaptive variation but far less than in the native range and we discuss the implications for the predicted level of virulence in Europe over the long term.

Hymenoscyphus fraxineus version 2 genome approaches chromosome level contiguity

The *Hymenoscyphus fraxineus version 2* genome assembly (*Hf*-v2.0) is 62.28 Mbp distributed across 145 scaffolds. Here, we analyse scaffolds greater than 10 Kbp, a total of 23 scaffolds or 62.08 Mbp (Fig. 2; Supplementary Analysis 2). Both *Hf*-v1.116 and *Hf*-v2 were released (open resources) online and hosted on the OADB github repository (https://github.com/ash-dieback-crowdsource). The *Hf*-v2.0 gene annotation pipeline built on that of *Hf*-v1.116 and identified 10,945 genes (11,097 transcripts, 1516.50 mean CDS length, 3.49

mean exons per gene; Supplementary Analysis 3). Telomeric repeats are identified at both ends of 14 scaffolds and at one end of all but one of the remaining scaffolds (Supplementary Table 2). Genetic linkage between scaffolds in the progeny of a *H. fraxineus* lab cross suggests linkage between three pairs of scaffolds missing a telomere (Supplementary analysis 4). By joining these linked scaffolds, we estimate 20 chromosomes (pseudomolecules) for *H. fraxineus* (Fig. 2 track b).

The European invasion population is bottlenecked and sexually recombining diversity from Asia

In 43 European and 15 Japanese haplotypes, we identified 6.26 million single nucleotide polymorphisms (SNPs) overall (SNPs segregating in: Japan (Jp) = 4.5×10^6, Europe (EU) = 0.67×10^6; Supplementary analysis 5). A SNP network of all genes shows us that the European population is genetically divergent and bottlenecked from the native population (Fig. 3). The disparity in nucleotide diversity (π) between Europe and Japan (Fig. 3) could have resulted from a founding population much smaller than that of Japan. Tajima's *D*[17] is a statistic, centred around zero, that is sensitive to changes in effective population size (and/or the mode of selection). In Japan, we observe a Tajima's *D* value close to zero ($\bar{x} = -0.22$) which indicates neutrality with purifying selection operating on genes ($\bar{x} = -0.89$; Fig. 3). In Europe, the genomic signal is broadly positive ($\bar{x} = 1.28$; Fig. 3). This positive value is generated by a contraction in population size balancing allele frequencies by reducing the frequency of common alleles without equivalent reduction in the frequency of rarer alleles (see Supplementary Fig. 10). This balancing of allele frequencies without a real reduction in SNP density is consistent with a small founder EU population from a larger source.

Sex is a fundamental determinant of plant pathogen adaptive potential not least because it allows recombination of host interaction genes, or effectors[18]. Here, as well as observing sexual reproduction between European isolates in the lab (Supplementary analysis 4) we see a breakdown in linkage disequilibrium in the wild in Europe (Supplementary Analysis 6). At the genome level Japanese and European populations are divergent (F_{ST}=0.58 (95% CI=0.54-0.56)) but recombination decouples relationships between genes. Across the genome, genetic differentiation correlates positively with diversity present in the Japanese population (R^2=0.18, n=627, p<0.001) but negatively with diversity present in the European population (R^2=0.51, n=610, p<0.001). That is to say that, the more diversity present at a locus in Japan, the less likely it is to be shared with Europe and, the more diversity present at a locus in Europe the more likely it is to be shared with Japan; a signal consistent with directionality from east to west (Fig. 2, tracks f & g, Supplementary Fig. 11). Preventing continued gene flow from the native range into Europe is important because sexual reproduction between divergent demes may be important for plant pathogen adaptation[19].

The European population was founded by two individuals from a large diverse population

Burokiene *et al.*[20] reported similar levels of genetic diversity (based on 11 microsatellite loci) between central European populations and those on the epidemic disease front. The authors suggest that for introduced diversity to reach the epidemic disease front, either the centre of genetic diversity quickly recombined and spread by range expansion, or only a small number of divergent *H. fraxineus* isolates arrived in Europe and present-day diversity

is the product of recombination among those. Here, we use third base positions of the coding regions of Core Eukaryotic Genes (CEGs) to distinguish between these scenarios and describe the invasion process. CEGs are highly conserved essential proteins, encoded in all eukaryotic genomes[21]. Their importance to fundamental eukaryotic biology makes the 387 CEGs we identified in all *H. fraxineus* individuals an ideal set to explore the European invasion. The bottleneck into Europe reduced the number of CEG haplotypes to 2.3 with 42% of all CEGs being monomorphic. The average number of Japanese haplotypes was 12.6 with no less than two per locus. Strikingly, European CEGs grouped into two divergent haplotypes. Furthermore, the level of divergence between those haplotypes reflects that across Japanese CEG networks (Fig. 4, Supplementary Fig. 20). It is our interpretation that these divergent haplotype pairs have been introduced by two haploid founders, as first suggested by Gross *et al.*[13]. Moreover, the divergence between these European haplotype pairs, without intermediates, represents the ancestral divergence of a large population from which the European population was founded.

The observed estimate for the haploid effective population size ($\vartheta_\pi = 2N_e\mu$) in Japan is 2.46 million individuals (π=0.0246) and the estimate from Europe is 0.67 million individuals (π=0.00672; μ=5×10^{-9} per base per generation). However, this European estimate of nucleotide diversity is inflated by the ancestral divergence of the two haploid founders. We used this divergent polymorphism to estimate the size of the European source population. Coalescent simulations of a hypothetical European source, bottlenecked to two individuals, show that as the source effective population size increases, so too does the average divergence between founder haplotypes. We find that a source effective population size between 2.2 and 2.8 million individuals (median 2.5M) most accurately replicates the observed European haplotype divergence (Fig. 5). Our estimate of the size of the source of the European *H. fraxineus* population is therefore equivalent to that from a single Japanese wood. These estimates may reflect the equilibrium diversity in any given *H. fraxineus* population within its native range, but importantly also suggests that the European invaders could have come from a single site, perhaps even from a single ascocarp (fruiting body).

10% of the H. fraxineus proteome putatively interacts with the host

Effectors are a broad classification of proteins that are secreted by bacteria, oomycetes, fungi, nematodes and aphids in order to interact with the host, disable host defence components and facilitate colonisation[15]. As such, effectors are in a co-evolutionary arms race with host resistance genes and are studied for insights into pathogen adaptive potential[22]. We used localisation signals of N-terminal presequences to identify secreted proteins in the *H. fraxineus* genome (Fig. 2, track e). These 1,132 predicted secreted proteins are potential effectors which clustered into 566 tribes (Supplementary Table 5). We did not identify any conserved fungal N-terminal signal motifs. A reduced set of 223 putative effectors were identified using a machine learning approach (Supplementary Analysis 7). Nevertheless, we consider the broadest set of secreted proteins in subsequent analyses to avoid missing out on potentially important effectors.

Putative effector genes are spread across the genome, away from repeat regions, similar to non-effector genes (Fig. 2 tracks c-e; Supplementary Fig. 21 & 22) which contrasts with

predictions of the two-speed genome model, in other filamentous plant pathogens23 (e.g. *Sclerotinia sclerotiorum*24). Absence of the two-speed evolution model has been observed in the genomic landscapes of members of the Magnaporthaceae family25. Perhaps, in *H. fraxineus,* the requirement for a dynamic process that operates to shuffle and generate novel effectors is lower in this sexually reproducing pathogen with high haplotype diversity26. The presence of sex and high haplotype diversity in turn indicate a long term balanced relationship between host and pathogen within the native range27

Structure and function of H. fraxineus putative effectors

Of the 1,132 predicted *H. fraxineus* secreted proteins, 62% carry Pfam domains associated with highly diverse biological functions (Supplementary Table 6). The largest subgroups include apoplastic catalytic enzymes, with 127 proteins predicted to have glycosyl hydrolase activity. Amongst these are putative cell wall degrading enzymes such as cellulases, pectinoesterases and cutinases. 77 predicted secreted proteins have oxidoreductase activity, 71 carry a cytochrome P450, 56 harbour conserved domains of unknown function and 396 did not carry a Pfam domain. To our knowledge, *H. fraxineus* is the first sequenced fungal pathogen with such high number of predicted secreted Cytochrome P450 proteins. Cytochrome P450s mode of action is typically via the monooxygenase reaction which is important in the generation and destruction of chemicals especially aromatics. In *H. fraxineus*, they may be important in pathogenesis, especially as they carry a signal of diversifying selection (Supplementary analysis 8). Potential roles for P450s in pathogenesis include: 1) destruction of ash tissue-derived aromatic compounds with antifungal properties. Plant secondary metabolism pathways are active during the infection process of *Magnaporthe oryzae*28. 2) Penetration of ash tree tissues through extracellular oxidation of wood and metabolism of hydrocarbon compounds which are the main constituents of host plants cuticle29. 3) P450s are part of the biosynthetic pathways for mycotoxins and phytohormones secreted by *H. fraxineus* during invasion of ash tissue, similar modes of actions have been shown for *Fusarium* multifunctional cytochrome P450 monooxygenase Tri430. The presence of such high number of P450s predicted in the secretome of *H. fraxineus* suggests biochemistry unique to *H. fraxineus*.

Transcribed low complexity domains may provide effectors with flexibility and diversity driven by adaptive evolution31. Of 1,132 effectors, 22% contain short repeats (e.g. Tetratricopeptide repeats and Leucine-rich repeats; Supplementary Table 7). Some are predicted to have a nuclear localisation signal often indicative of intracellular function in host cells whilst others have possible apoplastic function. Many putative effectors (37%), are cysteine-rich with between one and five predicted disulphide bonds (Supplementary Table 8) which are predicted to confer stability, biological activity and resistance to proteases in the apoplastic space32 and inside infected host cells33.

We identify three effectors (one of them previously identified16) with an NPP1 (necrosis-inducing Phytophthora protein) domain which is present in fungal, oomycete and bacterial proteins that induce hypersensitive reaction-like cell death upon infiltration into plant leaves34. Two other proteins have predicted cell death activity and two more carry S1/P1 nuclease activity domains, which are involved in non-specific cleavage of RNA and single

stranded DNA. Host cell death induction is suggested to be common among facultative parasites that engage in a necrotrophic lifestyle[35].

Adaptive effector diversity is present in Europe, but is far greater in the native range

We analysed genetic diversity in all the genes of the European and Japanese populations (Supplementary Table 9). First, there remains a significant positive correlation in the level of polymorphism present in genes between the invasive European and native Japanese populations, despite an 81% reduction in SNP density in Europe (Fig. 6). This reduction in the level of polymorphism has impacted putative effector and other genes alike, whereas, in Japan putative effector genes maintain significantly greater levels of polymorphism than that of other genes (Fig. 6).

Putative effectors in Europe have an increase in SNPs that affect splice regions as well as 5' UTRs. Importantly however, there are also increases in SNP densities in synonymous, intron, up- and downstream positions in putative effectors relative to other genes (Supplementary analysis 8). Therefore, increases within these effector features could be the product of increased SNP density through linkage and balancing selection[36] operating on effectors.

Pathogen effectors must adapt to avoid host recognition and so they are expected to undergo positive, diversifying, and balancing selection[15]. To investigate the role of selection in the maintenance of polymorphism we measured adaptive diversity using the mean pairwise ratio of non-synonymous (P_N) to synonymous (P_S) polymorphism present in all genes. The P_N:P_S ratio can detect the presence of adaptive evolution when applied to short regions of a gene, e.g. a binding domain, because a value greater than one indicates positive or diversifying selection (functional divergence)[37] and P_N:P_S close to zero indicates negative or purifying selection. Here, we apply the P_N:P_S ratio across whole genes which we presume have different modes of selection operating across them. However, we expect effectors will retain an overall higher P_N:P_S value despite dilution by purifying selection operating over the rest of the gene.

Approximately 97% of all P_N:P_S values in the European and Japanese effectors are lower than one. Within this range, we observe a significant increase in the effector P_N:P_S over that of other genes between 0.2-0.5 (Fig. 6). This represents a reduction in the strength of purifying selection in effectors relative to other genes. At P_N:P_S values greater than one, we observe a significant peak in the density for effectors which indicates a number of effectors evolving under contemporary positive selection (Fig. 6). Effectors in Japan have higher P_N:P_S values than other genes (Fig. 6). In Europe, effector adaptive diversity has been maintained despite the bottleneck, albeit at a lower level than in the native range (Fig. 6). Finally, consistent with linkage and balancing selection[36], the level of synonymous polymorphism is also higher in effectors in both Japan and Europe (supplementary Fig. 25). Pseudogenes too, may have a P_N:P_S ratio approaching one. Pseudogenisation could be more readily observed for effectors increasingly recognised by the immune system[38]. Here, we do not observe a significant increase in start loss, stop gain or stop loss variants in effectors over other genes (Supplementary analysis 8).

Conclusions

The bottleneck of *H. fraxineus* into Europe removed the majority of neutral and much of the adaptive genetic variation. Despite this, the pathogen has devastated ash from east to west and host defence is characterised in terms of levels of susceptibility5. Further introduction of pathogen genetic diversity runs the risk of increasing disease prevalence. Successive invasions can increase the level of genetic diversity above that of native populations39 and drive temporal fluctuations in adaptive diversity40. Increased virulence at invasion onset, due to uninfected host density, may be attenuated by natural selection on the pathogen as uninfected density reduces41 but this is much less probable where immigrations continues.

Tree pathogens can be introduced through trade in live trees and untreated wood1. If multiple native genotypes can invade, we face a situation where further immigration is likely to be accompanied by extreme increases in the level of adaptive polymorphism and disease severity. It is most important to prevent further introduction of pathogen diversity to Europe, particularly by prohibiting imports of *Fraxinus* species from East Asia, including material which may contain leaf debris bearing ascocarps.

The *H. fraxineus* native range, Asia, is large and the fungus lives there on the leaf litter of several ash species. Broader population genetic analyses over this range will allow us to address important questions on *H. fraxineus* adaptation to specific hosts. Moreover, the two-haploid founder scenario provides a unique system to unravel the population genetics of invasion with genome evolution. Finally, efforts to gain an understanding of disease progression will combine effector nucleotide statistics with targeted RNA-seq experiments.

Notwithstanding the potential impact of an introduction of the emerald ash borer42, current levels of European pathogen genetic diversity may infect and kill 95% of all European ash. We must consider the implications of any further introduction of diversity from the ash dieback pathogen to an already dire situation.

Materials & Methods

Hymenoscyphus fraxineus collection and sequencing

44 *Hymenoscyphus fraxineus* isolates were collected from Europe (21 UK, 13 Norway, 5 France, 4 Poland, 1 Austria) and nine from Japan (Supplementary Table 1). Japanese isolates were made up of one haploid and eight fruiting bodies (see isolate phasing below). Fruiting bodies were stored in ethanol and haploid isolates were cultured on malt extract agar43. Species confirmation was performed by PCR and sequencing of ITS sequences. Isolates were genome sequenced at either Edinburgh Genomics or The Earlham Institute on MiSeq and HiSeq Illumina platforms at 25-160 sequence depth (Supplementary analysis 1).

Genome Assembly and annotation of H. fraxineus

As part of a commitment to open science and rapid community analyses, the genome of isolate KW1, *Hf*-v1.1, was released online in March 2013 on the open ash dieback (oadb.tsl.ac.uk) repository16. *Hf*-v2, published here and released online in February 2015, was assembled using a 200bp insert Paired End (PE) and a Long (5 Kbp) Mate Pair (LMP)

library. SOAPec error correction was performed on PE reads (overlapping merged) and SOAPdenovo v2.0444 was used to assemble and scaffold (Supplementary Analysis 2).

The annotation pipeline used for genome *Hf*-v1.116 was replicated on *Hf*-v2 adding newer RNA-seq data from the Open Ash Dieback repository (Supplementary Analysis 3). The *Hf*-v2 genome was repeat masked using *de novo* models, known fungal and low complexity repeats. Annotations (9,737) from *Hf*-v1.1 were transferred to *Hf*-v2. Augustus v2.745 was used to predict additional protein-coding genes using protein and RNA-seq alignments (16 libraries, Supplementary Table 16) with the previously trained *Hymenoscyphus fraxineus* model. We ran a transcript extension protocol on all genes to identify whether those genes within 100bp of an upstream gene had evidence for extension. We also checked this extended gene dataset for potential secreted proteins. This protocol allowed the identification of 36 genes that had not previously been recognised as having a methionine start codon which were included into our effector-mining pipeline (Supplementary analysis 3).

We used the pipeline described by Saunders et al.32 to identify putative effectors. Briefly, transcripts with a signal peptide were identified and from these we remove transcripts with transmembrane or mitochondrial localisation signals and then transcripts with repeats, the disulphide bonds and nuclear localisation signals are identified. Finally, secreted proteins were clustered into tribes (see Supplementary analysis 8).

Mapping and SNP calling

Reads were trimmed for Illumina adapters and Phred quality (-q20) and read alignment was performed using BWA-MEM v0.7.746 after which duplicates were removed. VCFtools v0.1.1347 was used to filter variants to a minimum depth of ten, maximum depth of $1.8 \times$ mean individual depth and a SNP genotype quality of at least 30. SNP sites with more than two alleles were excluded as probable errors, as were sites reported as heterozygous in haploids. Finally, sites that were missing in 20% individuals were removed. Three samples were removed from further analyses because of insufficient depth or evidence of contamination, leaving 20 UK, 13 Norway, 5 France, 4 Poland, 1 Austria, and 8 Japan (1 haploid 7 fruiting bodies; see Supplementary Analysis 1).

Phasing fruiting body samples

We used shapeit248, to phase fruiting body data, this first applies a phase informative read step (extractPIRs) to group SNPs present on the same read pair and then phases remaining SNPs using population level polymorphism (assemble --states 1000 --burn 60 --prune 60 --main 300 --effective-size 564000 --window 0.5). All Japanese samples including the haploid strain (as control) were phased. As required indels were removed from the dataset before phasing. Effective population size ($\vartheta = 2N_e\mu$) was calculated assuming a mutation rate of 5×10^{-9} per generation per site49 using the mean nucleotide diversity present in 100 Kbp windows in the Japanese population.

SNP diversity and divergence analysis

Polymorphism statistics and sliding windows, π and F_{ST} were conducted using VCFTOOLS v0.1.1347. DNASP v5.10.0150 was used to calculate population genetic statistics per gene and PAML v4.951 was used (YN00) in pairwise mode to calculate the average $P_N:P_S$ ratio nonsynonymous (P_N) and synonymous (P_S) values for polymorphism in each gene with at least one synonymous mutation. The sm package52 in R v3.2.1 compared the density distributions of $P_N:P_S$. Output data for all genes are combined in Supplementary Table 9 (see also Supplementary analysis 6). Data are represented using SPLITSTREE v4.13.153 using the Neighbour net algorithm and CIRCOS v0.67-554.

Parental cross and linkage analysis

Two *H. fraxineus* isolates of opposite mating types, LWD054 and LWD067, were collected from stem lesions on trees in Norfolk, UK in 2014. Crosses were made according to the method of Wey *et al.*55 except that isolates were kept at 25°C in 16 hours light. KASP™56 markers were designed to SNPs between two *H. fraxineus* parents mapped to the *Hf*-v2 assembly (Supplementary Analysis 4). Linkage between scaffolds was ascertained using Chi-squared test of sites close to ends of the 23 scaffolds, using 28 F1 offspring. Scaffolds at linkage group ends were examined for the presence of telomeric motifs.

Statistics

R v3.2.157 was used to fit simple regression models to test the level of association between population signals in order to show directionality in the invasion and portray the population contraction (founder effect) behind Tajima's D. To evaluate the correlation in gene diversity (the proportion of SNPs per gene) in Japanese and the European populations we log-transformed SNP density excluding genes with less than two SNPs (leaving 6216 genes). SNP density regression models were calculated on log-log data and plotted on the original scale. We also correlated feature coverage (i.e. repeat content) of regions with gene and effector coverage to understand genome evolution. Two-sample Wilcoxon tests (one tailed indicated by '>' or '<'; two tailed indicated by '~'), were used to test for differences in signals of nucleotide diversity, Tajima's D and $P_N:P_S$ between non-effector and effector genes in European and Japanese populations. Bootstrap analyses compare variants effect by feature or gene (i.e. SnpEff output) and involved sampling feature with replacement (×1000). Bootstrapped comparisons were tested using a randomisation test and 95% confidence intervals (CI) are presented.

Ancestral effective population size

To estimate the size of the population from which the European population was founded we used the SNPs at the third base pair in codons within 387 core eukaryotic genes. CEGMA v2.421 initially identified 440 CEGs which was reduced to 387 genes that were confirmed by reciprocal blastp (BLAST v2.2.3158 (-evalue 1e-5) with a minimum of 90% coverage (removing technical assembly errors and biological duplicates), removing CEGs with nonsense variants (expected pseudogenes) and a minimum mean depth of 10x per gene across all individuals (poorly covered CEGs). Remaining genes were then run through a pipeline to quantify the pairwise distance between alleles (using SPLITSTREE) and the

diversity per gene (using DNASP v5). Relative pairwise distances between (third base pair) haplotypes is plotted as a density including all 387 CEGs combined. CEG networks are shown to as a visual demonstration of the shapes of three polymorphic networks in Japan and Europe HYMFR746836.2.0_000003240.1, -000004200.1, -000061770.1)

The average number of SNPs (3rd codon bp) per CEG was 448 and this number was used as an input to FASTSIMCOAL2 v2.5.2.859, a fast sequential markov coalescent simulator used to estimate the size of the population from which the founders into Europe came. A simple model (x1000) of a single (haploid) population of fixed effective size (N_e=100,000– 4,000,000) containing 387 unlinked CEGs that were freely mutating at a rate of $5×10^{-9}$ per base per generation49 was bottlenecked to two individuals. From these two individuals, we recorded the number of haplotypes and the haplotype divergence at each gene for comparison to observed data.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Ghelardini L, Pepori AL, Luchi N, Capretti P, Santini A. Drivers of emerging fungal diseases of forest trees. For Ecol Manage. 2016; 381:235–246.

2. Brasier CM, Buck KW. Rapid evolutionary changes in a globally invading fungal pathogen (Dutch elm disease). Biol Invasions. 2001; 3:223–233.

3. Drenkhan R, et al. Hymenoscyphus fraxineus is a leaf pathogen of local Fraxinus species in the Russian Far East. Plant Pathol. 2017; 66:490–500.

4. McKinney LV, Nielsen LR, Hansen JK, Kjær ED. Presence of natural genetic resistance in Fraxinus excelsior (Oleraceae) to Chalara fraxinea (Ascomycota): an emerging infectious disease. Heredity (Edinb). 2011; 106:788–797. [PubMed: 20823903]

5. Sollars ES, et al. Genome sequence and diversity of European ash trees threatened by ash dieback. Nature. 2017; :212–216. DOI: 10.1038/nature20786

6. Downie JA. Ash dieback epidemic in Europe: How can molecular technologies help? PLoS Pathog. 2017; 13:1–6.

7. Gross A, Zaffarano PL, Duo A, Grünig CR. Reproductive mode and life cycle of the ash dieback pathogen Hymenoscyphus pseudoalbidus. Fungal Genet Biol. 2012; 49:977–86. [PubMed: 23036580]

8. Fones HN, Mardon C, Gurr SJ. A role for the asexual spores in infection of Fraxinus excelsior by the ash-dieback fungus Hymenoscyphus fraxineus. Sci Rep. 2016; 6:34638. [PubMed: 27694963]

9. McDonald BA, Linde C. Pathogen population genetics, evolutionary potential, and durable resistance. Annu Rev Phytopathol. 2002; 40:349–379. [PubMed: 12147764]

10. Charlesworth B. Effective population size and patterns of molecular evolution and variation. Nat Rev Genet. 2009; 10:195–205. [PubMed: 19204717]

11. Gladieux P, et al. The population biology of fungal invasions. Mol Ecol. 2015; 24:1969–1986. [PubMed: 25469955]

12. Fry WE. *Phytophthora infestans*: New Tools (and Old Ones) Lead to New Understanding and Precision Management. Annu Rev Phytopathol. 2016; 54:529–547. [PubMed: 27359366]

13. Gross A, Hosoya T, Queloz V. Population structure of the invasive forest pathogen Hymenoscyphus pseudoalbidus. Mol Ecol. 2014; 23:2943–2960. [PubMed: 24819666]

14. Schoebel CN, Botella L, Lygis V, Rigling D. Population genetic analysis of a parasitic mycovirus to infer the invasion history of its fungal host. Int J Lab Hematol. 2016; 38:42–49. [PubMed: 26362346]

15. Rafiqi M, Ellis JG, Ludowici VA, Hardham AR, Dodds PN. Challenges and progress towards understanding the role of effectors in plant-fungal interactions. Curr Opin Plant Biol. 2012; 15:477–482. [PubMed: 22658704]

16. Saunders DG, et al. Crowdsourced analysis of ash and ash dieback through the Open Ash Dieback project: A year 1 report on datasets and analyses contributed by a self-organising community. bioRxiv. 2014; doi: 10.1101/004564

17. Tajima F. Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 1989; 123:585–595. [PubMed: 2513255]

18. Croll D, Lendenmann MH, Stewart E, McDonald BA. The Impact of Recombination Hotspots on Genome Evolution of a Fungal Plant Pathogen. Genetics. 2015; 201:1213–1228. [PubMed: 26392286]

19. McMullan M, et al. Evidence for suppression of immunity as a driver for genomic introgressions and host range expansion in races of Albugo candida, a generalist parasite. Elife. 2015; 4:1–24.

20. Burokiene D, et al. Genetic population structure of the invasive ash dieback pathogen Hymenoscyphus fraxineus in its expanding range. Biol Invasions. 2015; 17:2743–2756.

21. Parra G, Bradnam K, Korf I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics. 2007; 23:1061–7. [PubMed: 17332020]

22. Stukenbrock EH, McDonald BA. Population genetics of fungal and oomycete effectors involved in gene-for-gene interactions. Mol Plant Microbe Interact. 2009; 22:371–80. [PubMed: 19271952]

23. Raffaele S, Kamoun S. Genome evolution in filamentous plant pathogens: why bigger can be better. Nat Rev Micro. 2012; 10:417–430.

24. Derbyshire M, et al. The complete genome sequence of the phytopathogenic fungus Sclerotinia sclerotiorum reveals insights into the genome architecture of broad host range pathogens. Genome Biol Evol. 2017; 9:593–618.

25. Okagaki LH, Sailsbery JK, Eyre AW, Dean RA. Comparative genome analysis and genome evolution of members of the magnaporthaceae family of fungi. BMC Genomics. 2016; 17:135. [PubMed: 26911875]

26. Stukenbrock EH, Croll D. The evolving fungal genome. Fungal Biol Rev. 2014; 28:1–12.

27. Tellier A, Brown JKM. Stability of genetic polymorphism in host-parasite interactions. Proc R Soc B Biol Sci. 2007; 274:809–817.

28. Collemare J, et al. Magnaporthe grisea avirulence gene ACE1 belongs to an infection-specific gene cluster involved in secondary metabolism. New Phytol. 2008; 179:196–208. [PubMed: 18433432]

29. Syed K, Yadav J. P450 monooxygenases (P450ome) of the model white rot fungus Phanerochaete chrysosporium. Crit Rev Microbiol. 2012; 38:339–363. [PubMed: 22624627]

30. Tokai T, et al. Fusarium Tri4 encodes a key multifunctional cytochrome P450 monooxygenase for four consecutive oxygenation steps in trichothecene biosynthesis. Biochem Biophys Res Commun. 2007; 353:412–417. [PubMed: 17188234]

31. Mesarich CH, Bowen JK, Hamiaux C, Templeton MD. Repeat-containing protein effectors of plant-associated organisms. Front Plant Sci. 2015; 6:872. [PubMed: 26557126]

32. Saunders DG, et al. Using hierarchical clustering of secreted protein families to classify and rank candidate effectors of rust fungi. PLoS One. 2012; 7

33. Blondeau K, et al. Crystal structure of the effector AvrLm4-7 of Leptosphaeria maculans reveals insights into its translocation into plant cells and recognition by resistance proteins. Plant J. 2015; 83:610–624. [PubMed: 26082394]

34. Fellbrich G, et al. NPP1, a Phytophthora-associated trigger of plant defense in parsley and Arabidopsis. Plant J. 2002; 32:375–390. [PubMed: 12410815]

35. Gijzen M, Nürnberger T. Nep1-like proteins from plant pathogens: Recruitment and diversification of the NPP1 domain across taxa. Phytochemistry. 2006; 67:1800–1807. [PubMed: 16430931]

36. Hughes AL, Yeager M. Natural selection at major histocompatibility complex loci of vertebrates. Annu Rev Genet. 1998; 32:415–435. [PubMed: 9928486]

37. Stukenbrock EH, Bataillon T. A population genomics perspective on the emergence and adaptation of new plant pathogens in agro-ecosystems. PLoS Pathog. 2012; 8:e1002893. [PubMed: 23028308]

38. Burstein D, et al. Genomic analysis of 38 Legionella species identifies large and diverse effector repertoires. Nat Genet. 2016; 48:167–175. [PubMed: 26752266]

39. Uller T, Leimu R. Founder events predict changes in genetic diversity during human-mediated range expansions. Glob Chang Biol. 2011; 17:3478–3485.

40. McMullan M, van Oosterhout C. Inference of selection based on temporal genetic differentiation in the study of highly polymorphic multigene families. PLoS One. 2012; 7:e42119. [PubMed: 22900006]

41. Berngruber TW, Froissart R, Choisy M, Gandon S. Evolution of Virulence in Emerging Epidemics. PLoS Pathog. 2013; 9

42. Villari C, Herms DA, Whitehill JGA, Cipollini D, Bonello P. Progress and gaps in understanding mechanisms of ash tree resistance to emerald ash borer, a model for wood-boring insects that kill angiosperms. New Phytol. 2016; 209:63–79. [PubMed: 26268949]

43. Kowalski T. Chalara fraxinea sp. nov. associated with dieback of ash (Fraxinus excelsior) in Poland. For Pathol. 2006; 36:264–270.

44. Luo R, et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience. 2012; 1:18. [PubMed: 23587118]

45. Stanke M, Morgenstern B. AUGUSTUS: A web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. 2005; 33:465–467.

46. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv. 2013:1–3. 1303.3997v.

47. Danecek P, et al. The variant call format and VCFtools. Bioinformatics. 2011; 27:2156–2158. [PubMed: 21653522]

48. Delaneau O, Zagury J-F, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. Nat Methods. 2013; 10:5–6. [PubMed: 23269371]

49. Kasuga T, White TJ, Taylor JW. Estimation of nucleotide substitution rates in Eurotiomycete fungi. Mol Biol Evol. 2002; 19:2318–2324. [PubMed: 12446823]

50. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009; 25:1451–1452. [PubMed: 19346325]

51. Yang ZH. PAML 4: Phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007; 24:1586–1591. [PubMed: 17483113]

52. Bowman AW, Azzalini A. R package 'sm': nonparametric smoothing methods (version 2.2-5.4). 2014

53. Huson DH, Bryant D. Application of Phylogenetic Networks in Evolutionary Studies. Mol Biol Evol. 2006; 23:254–267. [PubMed: 16221896]

54. Krzywinski M, et al. Circos: an Information Aesthetic for Comparative Genomics. Genome Res. 2009; 19:1639–1645. [PubMed: 19541911]

55. Wey T, Schlegel M, Stroheker S, Gross A. MAT- gene structure and mating behavior of Hymenoscyphus fraxineus and Hymenoscyphus albidus. Fungal Genet Biol. 2016; 87:54–63. [PubMed: 26724599]

56. Semagn K, Babu R, Hearne S, Olsen M. Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): Overview of the technology and its application in crop improvement. Mol Breed. 2014; 33:1–14.

57. Team RC. R: A language and environment for statistical computing. 2017

58. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic Local Alignment Search Tool. J Mol Biol. 1990; 215:403–10. [PubMed: 2231712]

59. Excoffier L, Foll M. fastsimcoal: A continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. Bioinformatics. 2011; 27:1332–1334. [PubMed: 21398675]

**Figure 1.** *H. fraxineus* **in the wild in the UK.**
(a) Brown/orange lesions on leaves followed by (b) wilting. (c) Disease progression: lesions on shoots followed by diamond-shaped lesions on the stems. (d & e) Wild, sexual reproduction occurs annually on fallen leaf rachises in the leaf litter. (f) Dieback of the crown6.
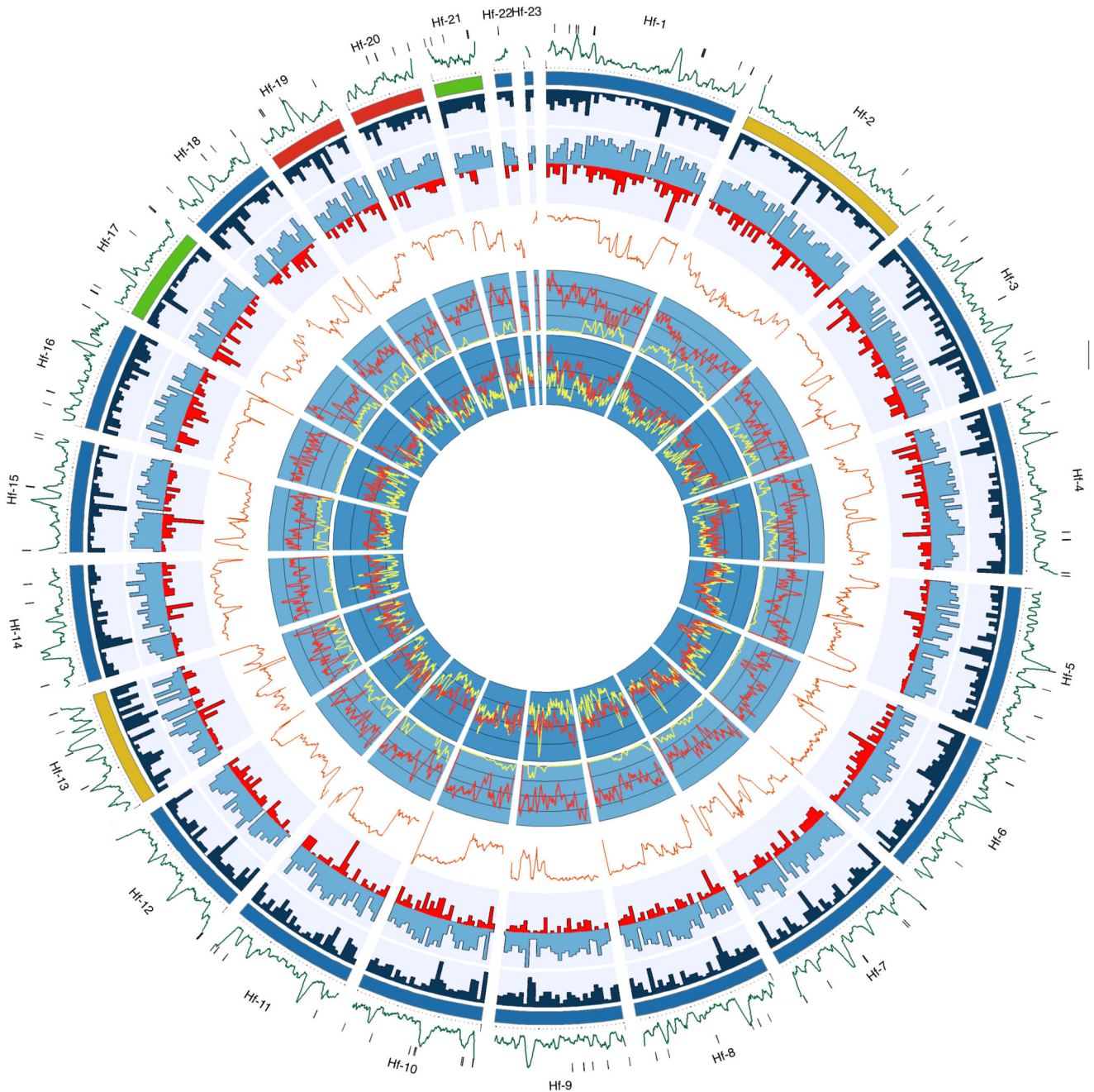
Figure 2. *H. fraxineus* genome organisation (scaffolds) and population statistics.
(a) AT richness sliding window (AT scale=0.5-0.85; window 63 Kbp by 5 Kbp slide) with outer ticks indicating gaps in the assembly >2.5 Kbp. (b) Scaffolds (ticks=100 Kbp) red, green and yellow pairs of scaffolds appear to be from the same linkage group. (c) Repeat density (0.0-1.0; 100 Kbp). (d & e) Gene (blue) & effector (red) density, respectively (0.0-1.0; 100 Kbp). (f) Genetic differentiation (F_{ST}) sliding window between Japanese and European populations (0.01-0.90; 100 Kbp by 5 Kbp). (g) Nucleotide diversity (π) sliding window within Japanese (red) and European (yellow) populations (0.0-0.032; 100 Kbp by 5

Kbp). (h) SNP density sliding window within Japanese (red) and European (yellow) populations (0.0-0.17; 100 Kbp by 5 Kbp). Key shows track reference and orientation.
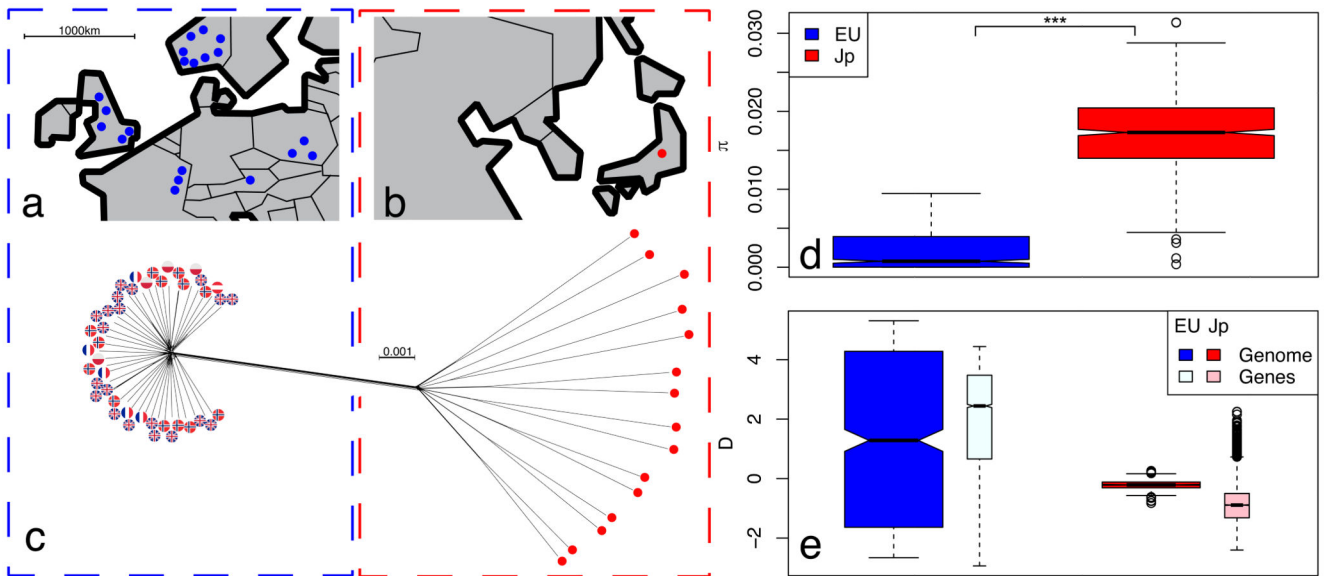
Figure 3. *H. fraxineus* European and Japanese sample sites and population genetic diversity.
(a) European and (b) Japanese sampling locations (to scale). (c) A neighbour net network of the concatenated coding regions of all genes in the genome (16.6 Mbp) for all European (n=43) and Japanese haplotypes (n=15) bounded by dotted line indicating the sampling location in panels a & b. (d) Boxplot of nucleotide diversity (π per 100 Kbp) for 642 windows in Japanese (Jp) and European (EU) populations. Nucleotide diversity is significantly greater in Japan than in Europe (Wilcoxon (π), Jp genome > EU genome: W(1257)=393210, n=1258, p<0.001[(***)]). (e) Boxplot of Tajima's *D* (*D* per 100 Kbp) for 642 windows across the genome and for each polymorphic gene within Japanese (n=10,869) and European (n=6,559) populations. In Japan, a signal of neutrality is present across the genome and tends to be negative in genes (purifying selection). The European population has a much broader positive distribution, consistent with a recent population decline. Boxplots show the median, upper and lower quartiles with outliers plotted outside whiskers which extend 1.5 times beyond the interquartile range.
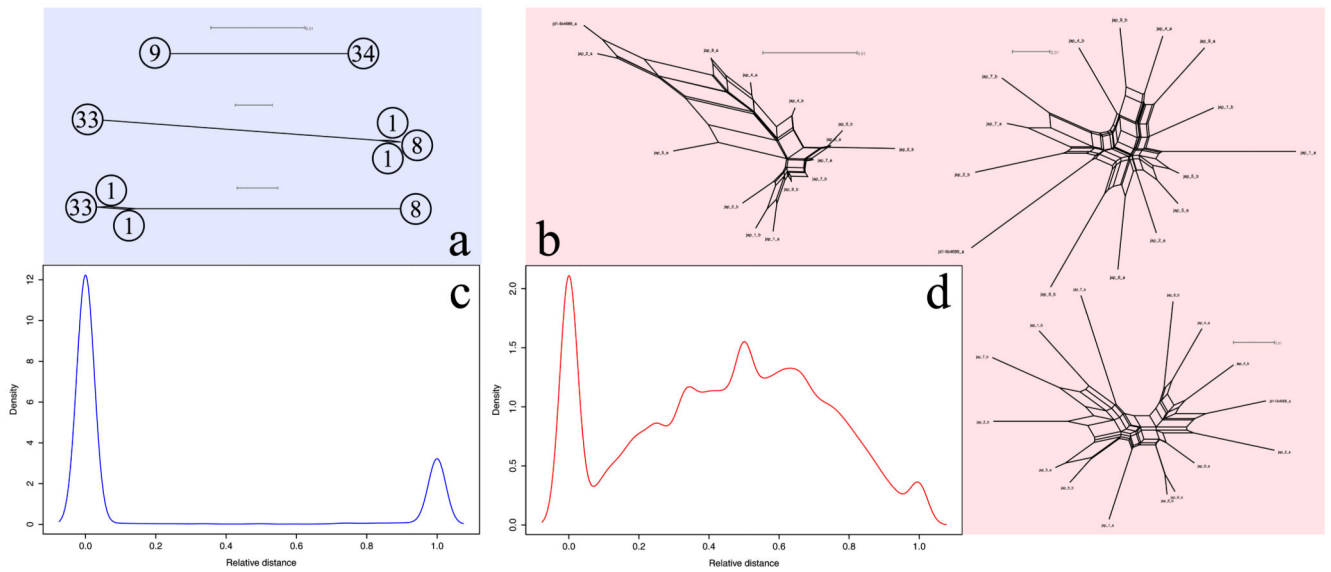
Figure 4. Divergence amongst core gene haplotypes in Europe is high and bimodal.
Neighbour net networks were generated individually for the third base position of 387 CEGs. A selection of networks built using three individual core genes are shown separately from European (a) and Japanese *H. fraxineus* (b) populations (see Supplementary Fig. 20 for combined networks). Encircled numbers are used in (a) to indicate numbers of individuals that share each haplotype. The three gene networks from the European population in (a) are the same genes as those from the Japanese population in (b) and drawn to the same scale. Density plots show the relative pairwise distance amongst haplotypes of all 387 CEGs in Europe (c) and Japan (d). In Japan, pairwise distances between haplotypes range in their divergence (d). In Europe, the majority of genes are either identical or at the complete opposite ends of the divergence range (c). The high level of divergence that separates European haplotypes, which are shared by many individuals, is visible in the three gene networks (a) but also in the measure of pairwise haplotype divergence of all 387 CEGs (c). This represents the presence of two major divergent haplotype groups in Europe.
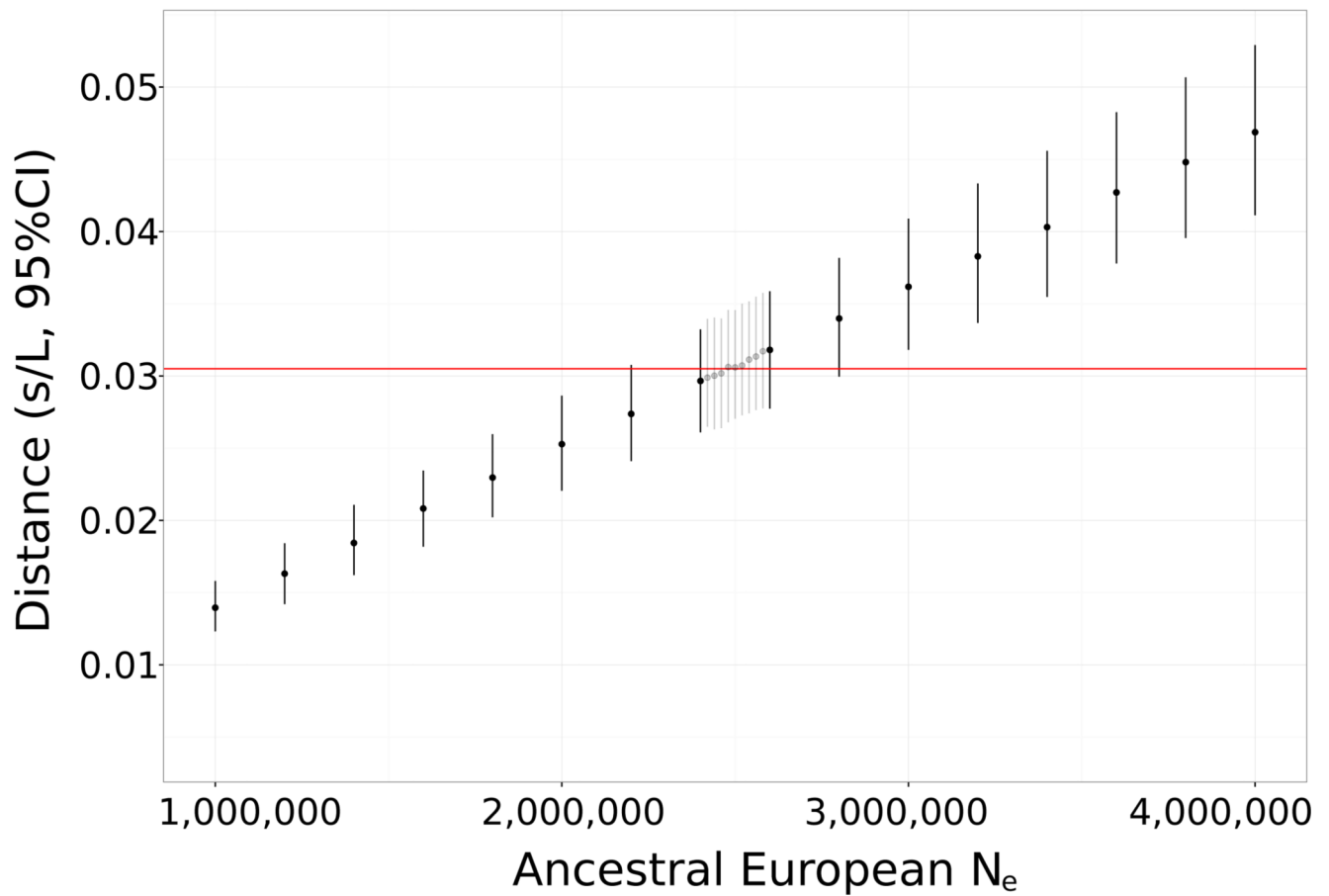
Figure 5. Coalescent simulations show a source population of 2.5 million individuals maintain enough polymorphism to account for the observed European haplotype divergence.
The coalescent of the third base pair of 387 core eukaryotic genes in a population was simulated (×1000) at an effective population size of between 1M and 4M haploid individuals. This (equilibrium) population was bottlenecked to two individuals and the diversity at all polymorphic loci was recorded. 2.5M individuals best accounted for the observed diversity in present day European divergent haplotypes (observed SNPs/bp=0.0305). Plot shows median and 95% confidence intervals.
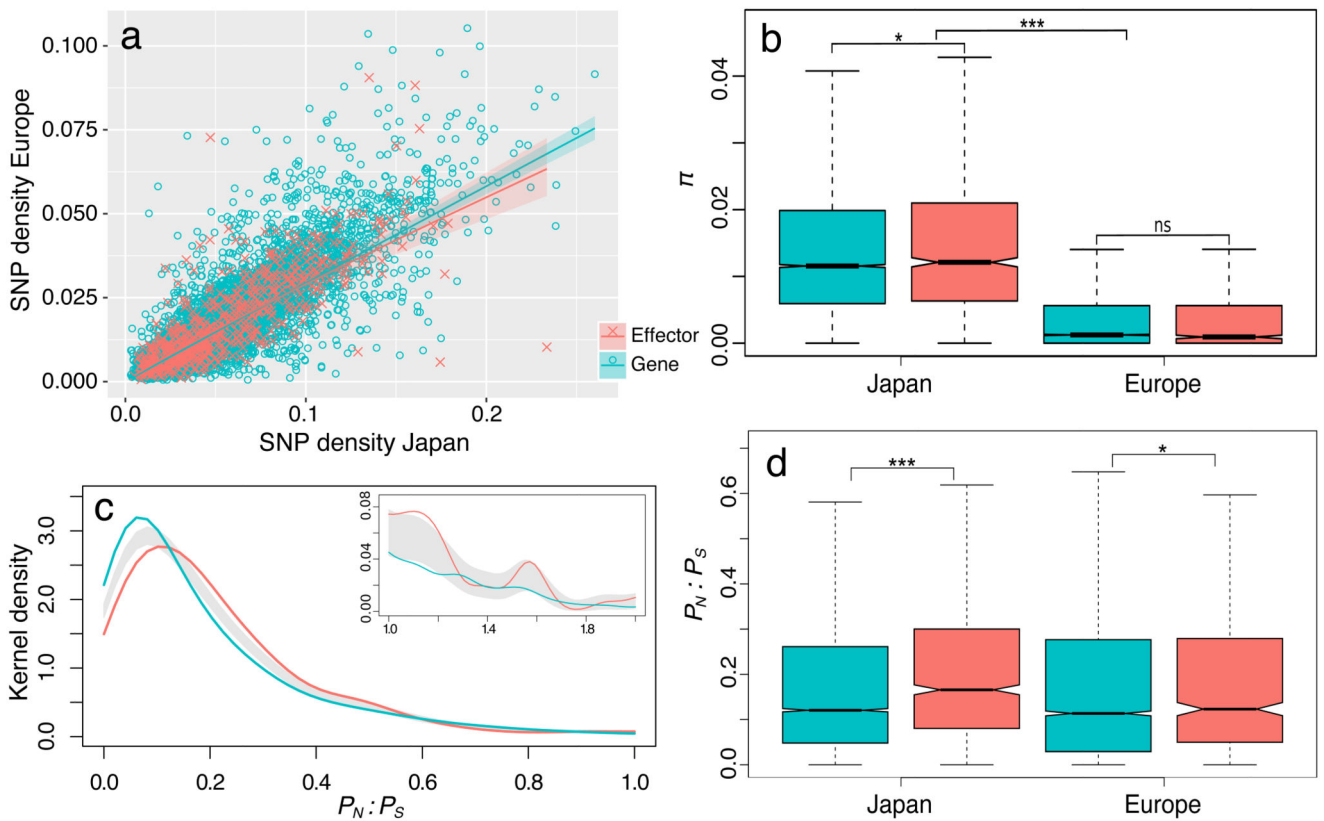
Figure 6. Neutral and adaptive genetic diversity in *H. fraxineus* genes and potential effectors.
(a) The Japanese population contains an increased level of diversity to that in Europe but despite the founder effect, genes that are more polymorphic in Japan tend to be more polymorphic in Europe (EU-Jp SNP density: gene, R^2=0.664, n=3214, p<0.001; effector, R^2=0.605, n=369, p<0.001). Linear model calculated using log-log data and plotted here on the original scale (see Supplementary Fig. 23 for log-log scale). (b) The level of genetic diversity in Europe is not significantly different between potential effectors (Wilcoxon (π), EU genes ~ EU effectors: W(10942)=5420600, n=10943, p=0.78[(ns)]) whereas, in Japan putative effectors carry significantly more genetic diversity (Wilcoxon (π), Jp genes < Jp effectors: W(10942)=5671200 n=10943, p=0.041[(*)]). Nucleotide diversity over all genes combined is significantly greater in Japan than in Europe (Wilcoxon (π), Jp all genes > EU all genes: W(21885)=102060000 n=21886, p<0.001[(***)]). (c) The signal of adaptive diversity (P_N:P_S) is significantly different between putative effectors and all other genes (permutation test of equality: n=9258, p<0.001; joint kernel density shaded). The increase in effector P_N:P_S to other genes, over this range, is evidence for reduced efficacy of purifying selection in this classification of genes. For genes with a P_N:P_S greater than one and less than two (c, inset) effector genes again peak outside the joint kernel density, consistent with the operation of positive selection driving adaptive change in these effectors (permutation test of equality: n=183, p=0.01). (d) The overall level of adaptive polymorphism is greater in effectors than other genes in both Japan and Europe (Wilcoxon (P_N:P_S), Jp genes < Jp effectors: W(9510)=4891800, n=9511, p<0.001[(***)]; EU genes<EU effectors: W(6080)=1750600, n=6081, p=0.045[(*)]). However, the strength of this signal has been

reduced in the European population. Boxplots show the median, upper and lower quartiles with outliers plotted outside whiskers which extend 1.5 times beyond the interquartile range (see supplementary Fig. 24 & 25 for outliers).