Molecular Plant Pathology

MOLECULAR PLANT PATHOLOGY (2017) 18(1), 75–89 DOI: 10.1111/mpp.12384

A gapless genome sequence of the fungus *Botrytis cinerea*

JAN A. L. VAN KAN[1,*,†], JOOST H. M. STASSEN[1,†,‡], ANDREAS MOSBACH[2], THEO A. J. VAN DER LEE[3], LUIGI FAINO[1], ANDREW D. FARMER[4,5], DIMITRIOS G. PAPASOTIRIOU[6], SHIGUO ZHOU[7], MICHAEL F. SEIDL[1], ELEANOR COTTAM[8], DOMINIQUE EDEL[2], MATTHIAS HAHN[9], DAVID C. SCHWARTZ[7], ROBERT A. DIETRICH[5], STEPHANIE WIDDISON[8] AND GABRIEL SCALLIET[2]

[1]*Laboratory of Phytopathology, Wageningen University, 6708 PB, Wageningen, the Netherlands*
[2]*Syngenta Crop Protection Münchwilen AG, Crop Protection Research, CH-4332, Stein, Switzerland*
[3]*Wageningen UR Plant Research International, 6708 PB, Wageningen, the Netherlands*
[4]*National Center for Genome Resources, Santa Fe, NM 87505, USA*
[5]*Syngenta Biotechnology Inc., Research Triangle Park, NC 27709, USA*
[6]*Syngenta Jealotts Hill International Research Centre, Bracknell, Berkshire RG42 6EY, UK*
[7]*Department of Chemistry, Laboratory of Genetics and Laboratory for Molecular and Computational Genomics, UW Biotechnology Center, University of Wisconsin, Madison, WI 53706, USA*
[8]*General Bioinformatics, Jealotts Hill International Research Centre, Bracknell, Berkshire RG42 6EY, UK*
[9]*Faculty of Biology, Technical University Kaiserslautern, 67653 Kaiserslautern, Germany*

SUMMARY

Following earlier incomplete and fragmented versions of a genome sequence for the grey mould *Botrytis cinerea*, a gapless, near-finished genome sequence for *B. cinerea* strain B05.10 is reported. The assembly comprised 18 chromosomes and was confirmed by an optical map and a genetic map based on approximately 75 000 single nucleotide polymorphism (SNP) markers. All chromosomes contained fully assembled centromeric regions, and 10 chromosomes had telomeres on both ends. The genetic map consisted of 4153 cM and a comparison of the genetic distances with the physical distances identified 40 recombination hotspots. The linkage map also identified two mutations, located in the previously described genes *Bos*1 and *Bcsdh*B, that conferred resistance to the fungicides boscalid and iprodione. The genome was predicted to encode 11 701 proteins. RNAseq data from >20 different samples were used to validate and improve gene models. Manual curation of chromosome 1 revealed interesting features, such as the occurrence of a dicistronic transcript and fully overlapping genes in opposite orientations, as well as many spliced antisense transcripts. Manual curation also revealed that the untranslated regions (UTRs) of genes can be complex and long, with many UTRs exceeding lengths of 1 kb and possessing multiple introns. Community annotation is in progress.

Keywords: genetic map, grey mould, optical map, SMRT sequencing.

* *Correspondence*: Email: jan.vankan@wur.nl
†These authors contributed equally to this work.
‡*Present address*: University of Sheffield, Department of Animal and Plant Sciences, S10 2TN Sheffield, UK

INTRODUCTION

The rapid development of sequencing methodology has resulted in an overwhelming increase in genome sequence information for many biological systems. Sanger sequencing methodology has been in vogue for over two decades (1985–2010), and was gradually replaced in the early 21st century by 454 and Illumina technologies, which have become increasingly affordable in the past decade. More recently, third-generation sequencing methods have been introduced and their data quality has increased with a concomitant price reduction. These developments have prompted biologists in many disciplines to sequence the genome of their favourite organism(s) *de novo*, or to increase the quality of existing data with the ultimate goal to close the existing gaps and obtain a truly finished genome sequence (Thomma *et al.*, 2016).

Fungi act as saprotrophic biomass decomposers, pathogens and mutualistic symbionts, and numerous fungal species are of medical, agricultural or industrial relevance. Fungi are attractive organisms to study because of their haploidy (in particular the phylum Ascomycota), genetic tractability and generally small genome size, associated with limited repeat content. To further facilitate research into fungal biology and evolution, an initiative to sequence 1000 genomes from species across the fungal tree of life is currently ongoing (Spatafora *et al.*, 2013).

Botrytis cinerea is a necrotrophic plant-pathogenic fungus from the class Leotiomycetes, family Sclerotiniaceae, which infects at least 1400 plant species (Elad *et al.*, 2015), including many crops of economic importance (Dean *et al.*, 2012; Williamson *et al.*, 2007). The fungus not only exploits the plant programmed cell death machinery to cause infection on vegetative host tissues, but also affects flowers and fruit (van Kan, 2006; Williamson *et al.*, 2007). There is recent evidence that *B. cinerea* can also

Table 1 Comparison of the optical map with the sequence assembly.

Chromosome	Optical length (kb)	Optical coverage	Average optical fragment size (kb)	Assembly length (kb)	5′-telomere	3′-telomere	Comment
1	4068	107	12.3	4109	✓*	✓	
2	3264	97	12.2	3341	✓	✓	
3	3208	90	12.3	3227	✓	✓	
4	3055	138	10.3	2472	580†	✓	rDNA repeat
5	2965	107	13.6	2959	50†	22†	
6	2657	91	13.9	2726	✓	✓	
7	2632	103	12.7	2652	28†	✓	
8	2617	92	13.7	2617	✓	✓	
9	2516	88	13.5	2548	25†	✓	
10	2421	91	12.5	2419	✓	42†	
11	2327	83	13.1	2360	25†	✓	
12	2322	101	12.8	2353	✓	✓	
13	2233	110	13.1	2258	✓	✓	
14	2156	117	12.6	2138	32†	✓	
15	2015	81	12.3	2028	✓	✓	
16	1967	97	12.4	1970	✓	43†	
17	250	12	22.7	247	✓	✓	
18	210	32	17.5	209	✓	✓	

*Telomeric repeat present in assembly.

†Approximate length (in kb) of telomeric sequence missing from the assembly, as estimated from the optical map.

systemically colonize plants without causing disease symptoms (van Kan *et al.*, 2014; Sowley *et al.*, 2009). Disease control by fungicides is common practice, but there is increasing concern about fungicide resistance development in the *Botrytis* population (Hahn, 2014; Leroch *et al.*, 2013). *Botrytis cinerea* was ranked as the second most important plant-pathogenic fungus, based on scientific and economic significance (Dean *et al.*, 2012), and has become one of the most extensively studied necrotrophic plant pathogens. There are many studies on the functional roles of genes in virulence, development, metabolism, signalling and fungicide resistance. Such studies rely on accurate gene models and expression information, and are strongly facilitated by the availability of genome sequences.

First draft versions of the genome sequences of two *B. cinerea* strains were based on Sanger sequencing data with only approximately 5–10× coverage (Amselem *et al.*, 2011). In spite of the extensive biological information gained from detailed annotation (Amselem *et al.*, 2011), the incompleteness of the genome sequence and the inaccuracy of gene models merited the generation of improved draft versions based on Illumina data (Staats and van Kan, 2012). However, the growing *B. cinerea* research community, as well as the fungal community and industrial parties involved in the development of grey mould control measures, would benefit from access to a finished genome for at least one strain that can serve as a reference for future research. We report here the use of novel sequence data, partially derived from third-generation sequencing platforms, an optical map and a genetic map, to generate a gapless, near-finished genome assembly of *B. cinerea* isolate B05.10.

RESULTS

Genome sequences that are derived only from the assembly of sequence reads potentially suffer from errors, especially around regions with repetitive sequences. This is also true for earlier versions of *B. cinerea* genome assemblies (Amselem *et al.*, 2011; Staats and van Kan, 2012). In addition to using novel sequencing data, we employed physical and genetic maps to provide support for *de novo* genome assemblies.

Optical map of strain B05.10

An optical map of *B. cinerea* strain B05.10 was constructed using the restriction enzyme *Bst*EII. The finished optical map comprises 18 chromosomes (Fig. S1, see Supporting Information) with a length of 42.9 Mb and an average 90× genomic coverage (Table 1). Two small chromosomes (250 and 210 kb) showed lower coverage, possibly as a result of the parameters set for optical mapping (optical contig size cut-off of 200 kb, in combination with a small number of *Bst*EII restriction sites, leading to larger than normal optical fragment sizes).

Genetic linkage map

Botrytis cinerea field isolate 09Bc11, carrying resistance to several fungicides, was sequenced with Illumina methodology, yielding 50× genomic coverage. Mapping reads on the improved draft (v2) genome of B05.10 (Staats and van Kan, 2012) identified >200 000 single nucleotide polymorphisms (SNPs) between B05.10 and 09Bc11. A cross was performed between 09Bc11 (as maternal parent) and B05.10, and ascospores were collected. Germlings of 70

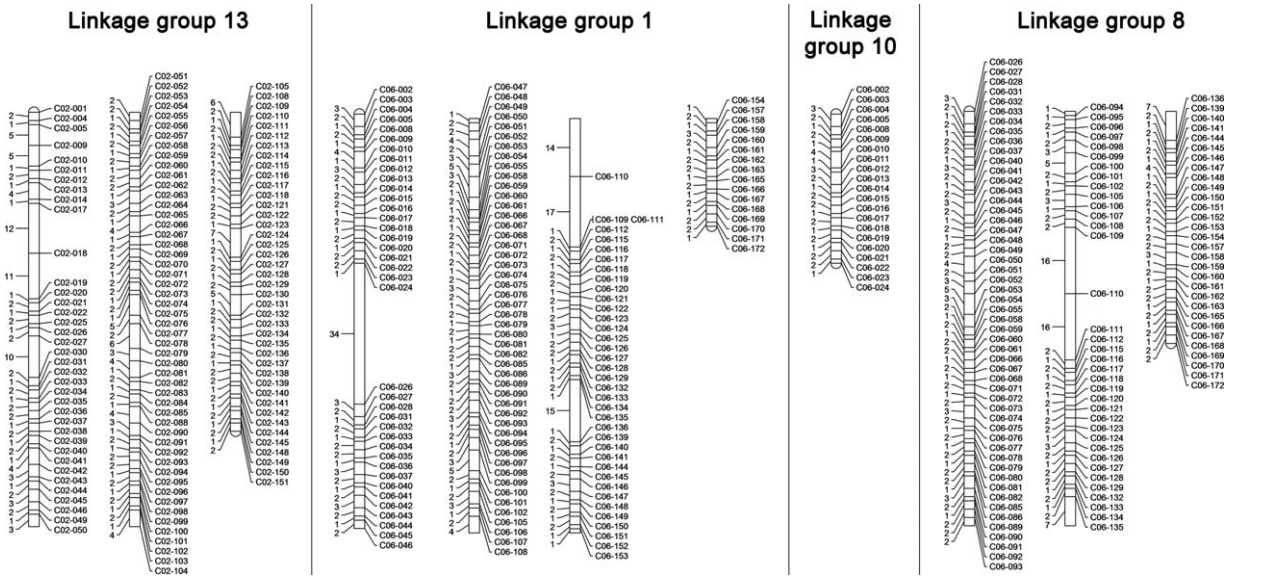

Fig. 1 Linkage map for a subset of the 18 identified linkage groups (LGs). LG13 (chromosome 2, Chr2) shows normal linkage patterns. The top arm of LG1 (Chr6) is identical to LG10, whereas the rest of LG1 is identical to LG8, indicative of a translocation.

ascospore progeny were cultured and their DNA was isolated and sequenced individually at approximately 40× coverage. Analysis of the segregation of SNP alleles revealed that three progeny had identical alleles to one of the parents, probably resulting from the contamination of asexual conidia in the ascospore suspension; data from these non-recombinants were eliminated from linkage analyses. Two recombinant progeny had identical allele patterns; one was eliminated from linkage analyses.

Filtering of SNPs yielded 2062 marker groups, which were used to calculate a genetic map with 18 linkage groups. Fifteen linkage groups [including LG13, = chromosome 2 (Chr2); Fig. 1] were as expected. The map, however, provided evidence for a translocation: LG1 (Chr6) appeared to be a merging of LG10 and LG8, with a genetic distance of 34 cM between markers C06-024 (the last marker on LG10) and C06-26 (the first marker on LG8). These markers were located on the same contig in the improved draft (v2) assembly of strain B05.10 (Staats and van Kan, 2012) at a distance of 43 kb. These observations are consistent with the existence of a translocation in isolate 09Bc11 between markers C06-024 and C06-026. The other breakpoint of the translocation was on LG11 (Chr1).

Resequencing strain B05.10

Strain B05.10 has been sequenced previously using Sanger (v1; Amselem *et al.*, 2011) and Illumina (v2; Staats and van Kan, 2012) methodology. In order to obtain higher coverage and better scaffolding, the Illumina read data (175-nucleotide paired-end and 3.5-kb mate pair libraries) used by Staats and van Kan (2012) were combined with new 10-kb mate pair reads and assembled. The assembly yielded 128 scaffolds with a length of 42 Mb, with an N50 of 2.44 Mb. Comparison of the genetic map with the order of marker groups on the assembly identified incongruences in the assembly. Such regions were manually split, reassembled and validated by inspecting the mapping of paired-end reads over the merged region. Gap closure was performed using previously available Sanger reads (Amselem *et al.*, 2011). The manually adjusted assembly of 41.5 Mb still had approximately 100 gaps (>20 nucleotides) and approximately 1000 ambiguous base calls in low-repetitive regions. In order to further close gaps and include telomeric regions in the assembly, single-molecule real-time (SMRT) sequencing and a *de novo* assembly were performed and yielded 123 contigs, which could be merged and integrated into 18 chromosomes, based on comparison with the optical map and alignment with the above Illumina-based assembly. Nine short contigs (<50 kb) containing telomeric repeats on one end remained unassigned to chromosomes, because they did not contain sufficient unique sequences to be connected to specific subtelomeric regions. The rDNA repeat appeared to be incorrectly assembled and was manually adjusted.

Mapping Illumina reads (from RNA and genomic DNA) on the PacBio consensus assembly revealed approximately 450 indels generated by the assembly software. On correction of the indels, the final assembly consisted of 18 chromosomes lacking internal gaps, with a total length of 42.6 Mb (containing a single copy of the rDNA unit at the start of Chr4). Chr17 and Chr18 are very small, 247.2 and 208.8 kb, respectively.

To ascertain that the assembly did not contain erroneously joint contigs, an *in silico* digest of the assembly was aligned to the optical map (Fig. S1). Figure 2 shows a summary of the results for three chromosomes that are representative of the dataset. In ten

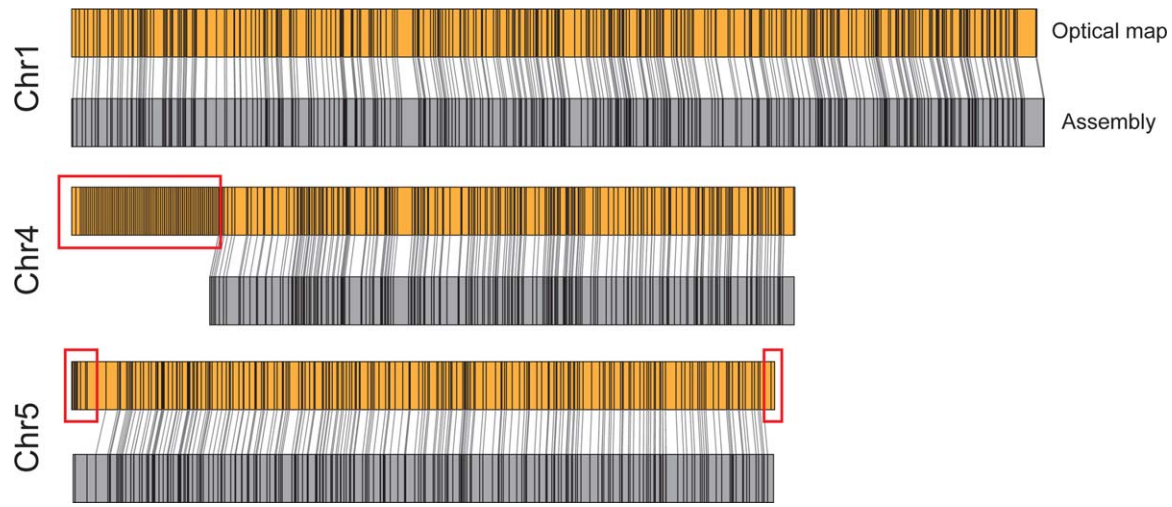

Fig. 2 Examples of alignments between the optical map (orange, top) and the *in silico* digest of the sequence assembly (grey, bottom) for three selected chromosomes. Each vertical line reflects the presence of a *Bst*EII restriction site, either experimentally determined (optical map) or predicted from the assembly. The red boxes highlight regions in the optical map which are not represented in the sequence assembly.

chromosomes (including Chr1), the optical map was congruent with the assembly over their entire length, including telomeric regions. The optical map of Chr4 was 580 kb longer than the assembly at its 5′-end. The rDNA repeat is present in a single unit in the assembly; however, the repetitive restriction pattern in the optical map is indicative of additional rDNA copies at the beginning of Chr4. The optical map of Chr5 was in agreement with the assembly, except for the absence of telomeric repeats at both ends (Fig. 2). Six other chromosomes contained a telomeric repeat at one end, but lacked one at the other end. The estimated lengths of sequences missing from the assembly at telomeres range between 22 and 50 kb (Table 1), bringing the length of missing sequence to 847 kb (267 kb for telomeric regions and 580 kb for rDNA repeats).

From the lack of internal gaps, we infer that all chromosomes contain fully assembled centromeric regions. Assuming that centromeres in filamentous fungi contain >20 kb of sequence, are devoid of genes and remain untranscribed (Smith *et al.*, 2012), the positions of centromeres can be proposed for 17 chromosomes (Table S1, see Supporting Information). Chr1 has two regions that might fulfil the criteria for centromeres (Fig. S2, see Supporting Information).

Analysis of meiotic progeny

The sequence reads of progeny from the cross between strains 09Bc11 and B05.10 were mapped onto the assembly. Read coverage over entire chromosomes was normalized and plotted for each individual (Fig. 3). For Chr1–16, the reads per kilobase per million (RPKM) values per chromosome were around 24 in nearly all cases, with two exceptions. First, for Chr4, the RPKM value was around 30 for all progeny. This was caused by the single copy of the rDNA repeat at the 5′-end of Chr4 (not shown), on which many mapped reads collapsed. Second, one single individual among the progeny (NGS0171-046L10) showed an RPKM value of approximately 30 for Chr1, whereas the RPKM value for other chromosomes in this progeny was normal (∼24). Further exploration of the read mapping depicting the coverage over Chr1 in this individual did not provide an explanation for this observation. Chr18 of strain B05.10 was lacking entirely in strain 09Bc11; hence, half of the progeny from the cross between these isolates did not contain any sequences derived from Chr18. Remarkably, however, the coverage of sequences derived from Chr17 and Chr18 varied between progeny and was often much lower when compared with the read coverage for 'core' chromosomes Chr1–16. For Chr17 and Chr18, some progeny showed RPKM values of five, whereas other progeny had values around 20 (Fig. 3).

Mapping of fungicide resistance loci

The strains used in the cross, 09Bc11 and B05.10, differ in fungicide sensitivity patterns. Strain 09Bc11 is resistant to boscalid and iprodione, whereas B05.10 is sensitive to both fungicides. To evaluate the mapping resolution achievable by the full progeny sequencing approach for independent loci, 67 sexual progeny of the cross between 09Bc11 and B05.10 were phenotyped on boscalid and iprodione. Resistant alleles segregated in a Mendelian 1 : 1 manner in the offspring (30R and 37S to iprodione; 37R and 30S to boscalid). Linkage was determined between fungicide resistance traits and SNP alleles (Fig. 4). Boscalid resistance fully co-segregated (at a frequency of >97%) with a single non-synonymous SNP allele that mapped to Chr1 around position 1 800 000, whereas iprodione resistance co-segregated (at a frequency of >97%) with 16 non-

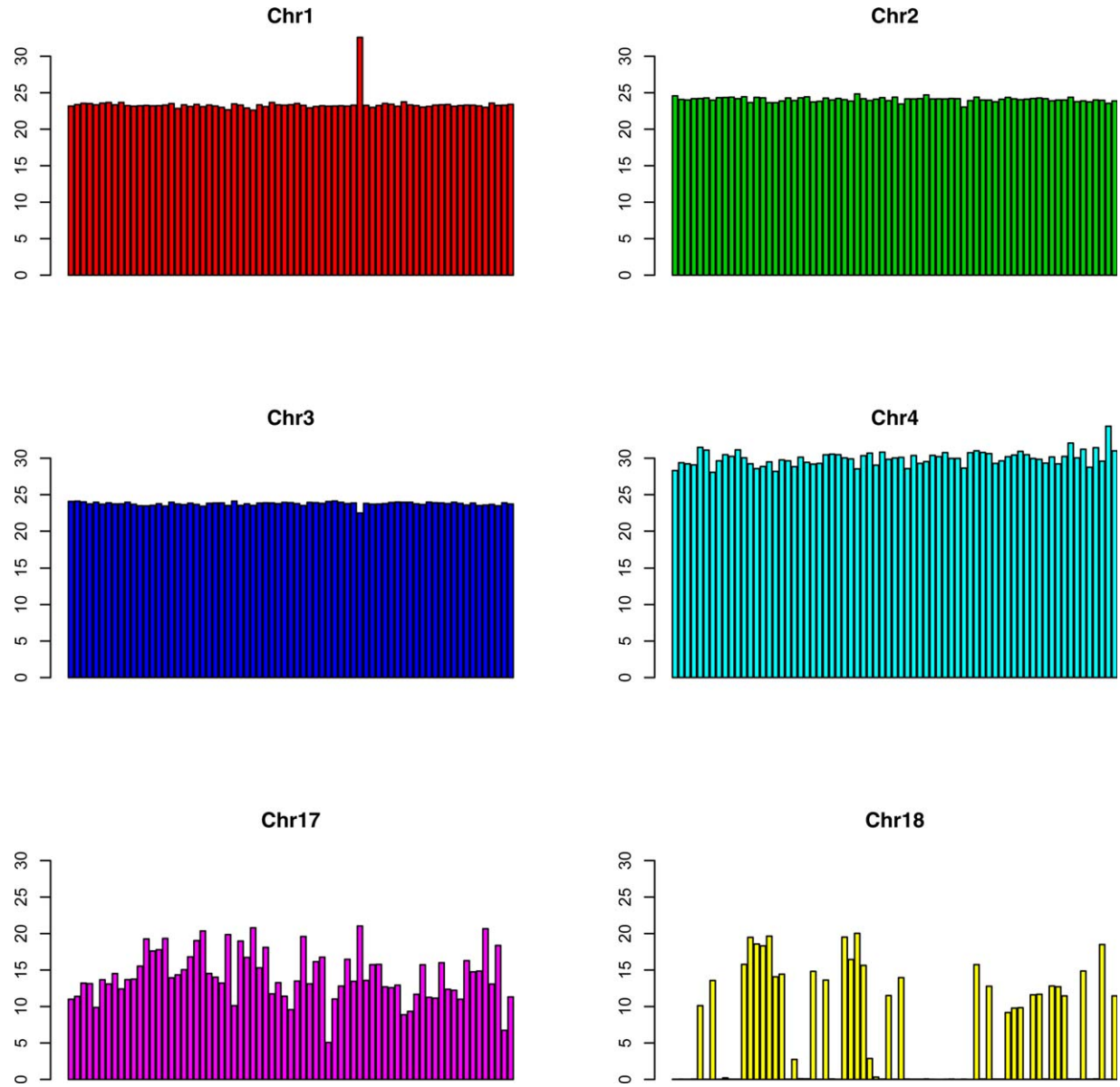

Fig. 3 Reads per kilobase per million (RPKM) values for a subset of chromosomes in sexual progeny from the cross 09Bc11 × B05.10. Each panel depicts RPKM values (*y*-axis) over one entire chromosome (indicated by their number) in 70 individual progeny as well as parental isolate 09Bc11. Each isolate is represented by one vertical bar in the graph, with isolate 09Bc11 on the extreme right. Plots for chromosomes 5–16 (Chr5–16) look similar to those for Chr2 and Chr3.

synonymous SNP alleles on Chr1 around position 2 200 000. Further examination of these regions showed that boscalid and iprodione resistance were caused by the known H272R mutation in the succinate dehydrogenase gene *BcSdh*B (Leroux *et al.*, 2010) and the known I365S mutation in the histidine kinase gene *Bos*1 (Leroux *et al.*, 2002), respectively.

Recombination rates

Using the final assembly of strain B05.10 as a reference, SNPs between B05.10 and 09Bc11 were re-calculated. The average frequency of SNPs over the entire genome was 1 in 150 bp. The distribution of SNPs between B05.10 and 09Bc11 along the chromosomes was heterogeneous: there were very long regions with only few SNPs and other regions with a high density of SNPs (Fig. 5). Approximately 75 000 SNPs that could be mapped unambiguously were used to construct a high-confidence, high-density linkage map. The genetic map had a cumulative length of 4153 cM and the ratio of physical versus genetic distance ranged from 7.8 kb/cM (Chr6) to 11.6 kb/cM (Chr2), with a genome-wide average of 10 kb/cM. A genetic distance of 1 cM corresponded to a physical distance ranging from 0 to 138 kb. The linkage map identified 40

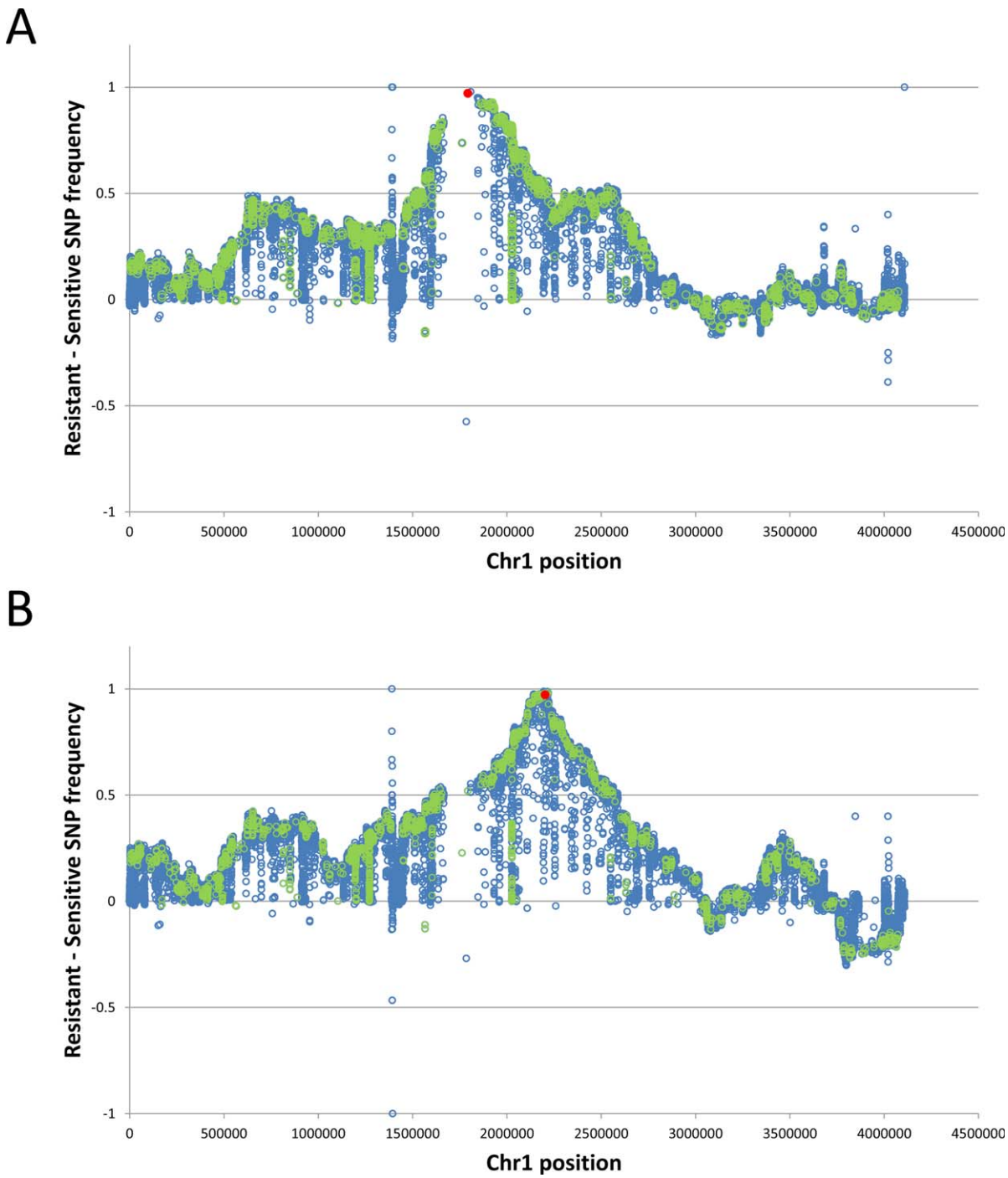

Fig. 4 Mapping of two fungicide resistance loci to chromosome 1 (Chr1). The *x*-axis represents sequence assembly coordinates. The *y*-axis represents the subtraction of frequencies at which a single nucleotide polymorphism (SNP) is found in respective fungicide resistant (R) minus sensitive (S) bulk. Frequencies were calculated for each position in the R or S bulk using the formula: total number of reads representing the SNP at position/total number of reads covering the position in bulk. The *y* value for any SNP position can vary between zero for unlinked positions to +1 for positively linked positions (SNP linked to resistance) or to −1 in the case of a negative linkage (SNP linked to sensitivity). Blue circles represent synonymous SNPs or SNPs outside coding sequences and green circles represent non-synonymous SNPs, exclusively. Red circles represent the SNPs that confer resistance. (A) Mapping of the boscalid resistance locus. (B) Mapping of the iprodione resistance locus.

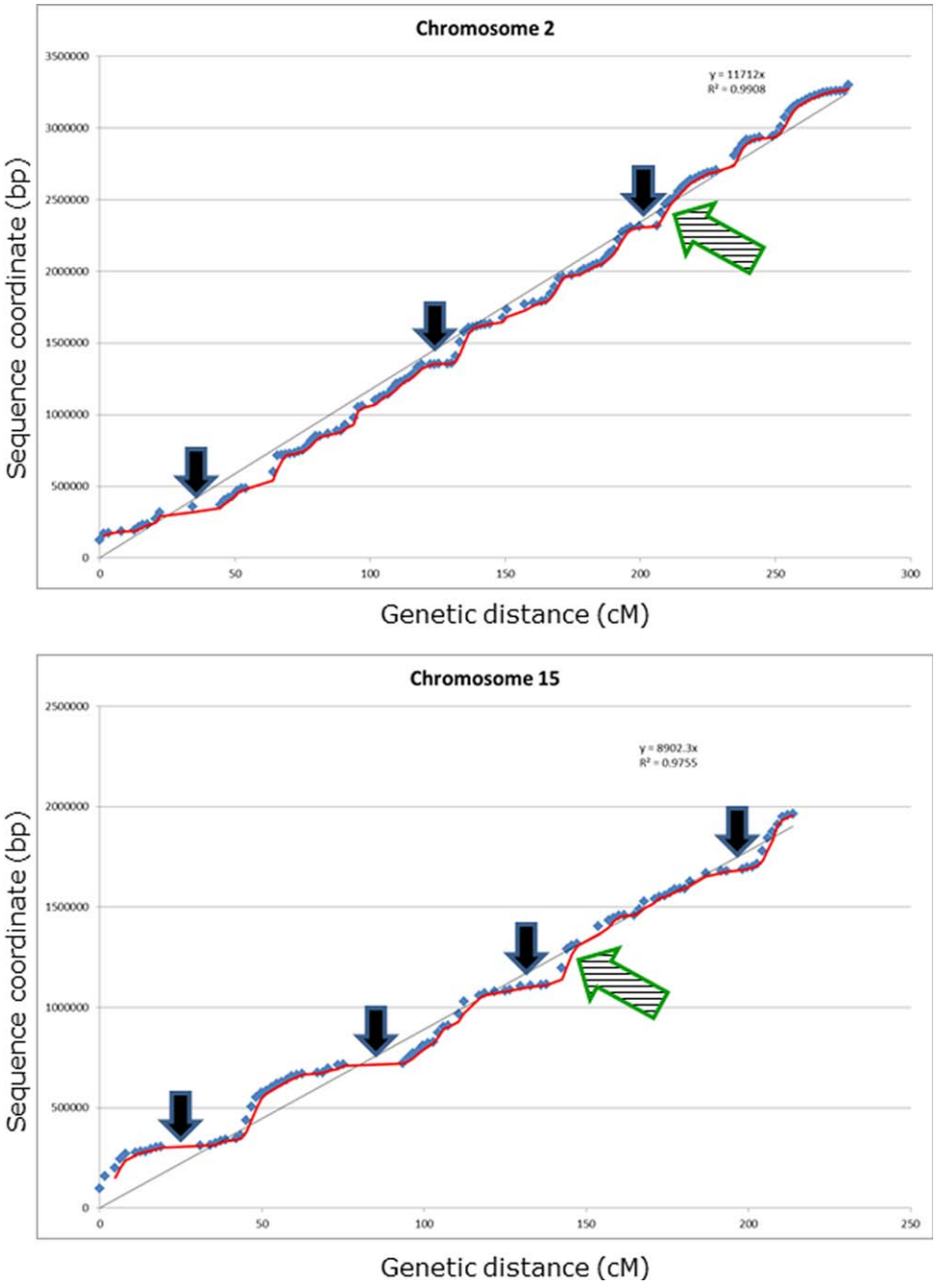

Fig. 5 Relationship between physical and genetic distances for chromosomes 2 and 15 (Chr2 and Chr15). The *x*-axis displays the genetic distance in centimorgan (cM) and the *y*-axis displays the chromosome coordinates in base pairs (bp). The correlation between physical and genetic distances and the correlation coefficient are given in the top right-hand corner of the graph. Each blue diamond represents a marker group [cluster of single nucleotide polymorphisms (SNPs) fully co-segregating in all progeny]. The green hatched arrow indicates the approximate position of the centromere. Black arrows indicate regions with high recombination rates.

regions (between one and four per chromosome) with high recombination rates (Figs 5 and S2). These regions were, with two exceptions, located at least 200 kb away from telomeric regions, whereas several recombination hotspot regions were adjacent to the centromeres, as in Chr2, 11, 14, 15 and 16 (Fig. S2). Regions with high recombination rates were not preferentially associated with the presence of transposable elements: only two of the 40 regions with high recombination rates contained a known transposable element (Table S2, see Supporting Information).

Gene prediction and annotation

Gene prediction was performed through an integration of tools, with support from intron hints derived from a pool of RNAseq data from biological samples representing conidia, mycelium, sclerotia, apothecia, ascospores and infected tomato tissues. Gene models corresponding to known transposable elements were identified by BLAST analysis with the transposons reported by Amselem *et al.* (2011), and were removed. There were, in total, 11 701 predicted genes. An estimate of the quality of the prediction was

obtained by manual curation of all gene models in Chr1, as well as a subset of gene models in other chromosomes. Of 1153 coding sequences analysed on Chr1, 1037 appeared to be entirely correctly predicted with respect to the translation start and termination sites, as well as splice junctions. The majority of errors (59) were caused by incorrect translation start site selection, whereas 41 gene models suffered from improper prediction of at least one (often non-canonical) splice junction. Such errors were manually corrected. A total of 181 genes on Chr1 possessed splice variants that represent a substantial proportion of the mRNA population for that specific gene.

Among the *B. cinerea* virulence factors that have been extensively studied are secondary metabolite gene clusters involved in the production of the phytotoxic compounds botrydial and botcinic acid. Strain B05.10 produces botrydial (Dalmais *et al.*, 2011; Pinedo *et al.*, 2008), yet neither the draft genome sequence (Amselem *et al.*, 2011) nor the improved draft (Staats and van Kan, 2012) contained an intact botrydial gene cluster. The gapless assembly contains an intact botrydial gene cluster (Bcin12g06370–Bcin12g06410), flanked by two repetitive regions of approximately 10 kb each. Furthermore, the botcinic acid biosynthetic genes, previously reported to be in separate clusters (Dalmais *et al.*, 2011), were located at the 5′-end of Chr1, with 16 kb of repetitive sequences separating the previously identified clusters. BcBOA1 is the first gene on Chr1 and is located only 5.5 kb from the telomeric repeat. Expression data suggest that the cluster contains only 13 genes (Bcin01g00010–Bcin01g00130, previously annotated as BcBOA1–13) based on co-expression in most samples (not shown). Gene models Bcin01g00140 and Bcin01g00150 (BcBOA15 and BcBOA16) are pseudogenes, whereas Bcin01g00160 (BcBOA17) has an expression profile very different from Bcin01g00010–Bcin01g00130.

Manual curation of Chr1 led to a number of further observations. There is one example of a transcript containing two predicted open reading frames (ORFs) that are translated from a single transcript. The ORF encoding the arginine-specific carbamoyl phosphate synthase small subunit (CarA, Bcin01g06930) is preceded by a short upstream ORF of 24 amino acids, which is a member of the Leader Carboxypeptidase A1 (CPA1) superfamily (described in *Saccharomyces cerevisiae, Neurospora crassa* and other filamentous fungi). These peptides are involved in the translational control of carbamoyl phosphate synthase biogenesis (Wang and Sachs, 1996; Wang *et al.*, 1998, 1999). In *B. cinerea*, there are 59 nucleotides between the stop codon of the leader peptide and the start codon of the carbamoyl phosphate synthase ORF.

Gene model Bcin01g02590 (encoding a protein of 100 amino acids) is entirely located inside a long intron in Bcin01g02580, and is supported by RNAseq. Gene model Bcin01g03590.3 has a non-canonical 3′-splice junction GT. . .AT, confirmed by RNAseq data, a complex splicing pattern and unusual coverage in the 5′-untranslated region (5′-UTR). Gene models Bcin01g04940 and Bcin01g04941, as well as Bcin01g06680 and Bcin01g06690, are overlapping over their entire length in opposite orientations.

Table 2 Statistics of the completeness of the *Botrytis cinerea* B05.10 genome based on 248 core eukaryotic genes (CEGs).

	No. of proteins	Completeness (%)	Total no.	Average	Ortho (%)
Complete	238	96.0	259	1.09	8.0
Group 1	62	93.9	68	1.10	6.4
Group 2	53	94.6	58	1.09	9.4
Group 3	59	96.7	64	1.08	8.5
Group 4	64	98.5	69	1.08	7.8
Partial	245	98.8	276	1.13	10.6
Group 1	65	98.5	71	1.09	6.2
Group 2	55	98.2	63	1.15	14.6
Group 3	60	98.4	68	1.13	11.7
Group 4	65	100	74	1.14	10.8

The lengths of 5′-UTRs of the gene models in Chr1 ranged from 26 to 1869 nucleotides, with a mean length of 290 nucleotides. The lengths of 3′-UTRs of the gene models in Chr1 ranged from 26 to 1446 nucleotides, with a mean length of 310 nucleotides. There are 14 genes on Chr1 which appear to have at least two introns in their 5′-UTR, and 25 genes with a 5′-UTR of at least 1000 nucleotides in length (Table S3, see Supporting Information). There are also 30 genes for which there is evidence for antisense spliced RNAs which overlap the coding sequence of the gene (Table S4, see Supporting Information).

Validation

The genome assembly of strain B05.10 and the quality of the proposed gene models were validated by several methods. First, alignments between optical contigs and *in silico* maps (Fig. S1) derived from the 18 assembled chromosomes were highly congruent: >99% of fragments aligned over their entire length, including putative centromeric regions. Second, the final assembly was totally congruent with the linkage map. Third, to assess the completeness of the assembled gene space, we searched for orthologues of core eukaryotic genes using the CEGMA pipeline and identified 98.8% of the CEGMA core genes in the final genome assembly as being completely or partially present (Table 2). The three KOGs that were not detected by CEGMA appeared to be present in *B. cinerea* and those genes were complete, suggesting a sensitivity problem of the CEGMA pipeline.

The quality of gene models was explored by comparing the identification of tryptic peptides in the predicted proteome of three different versions of the B05.10 genome. A sample of soluble proteins from mycelium was digested with trypsin, and the peptides were separated by iso-electric focusing and analysed by nano-liquid chromatography-mass spectrometry (NanoLC-MS). Peptide masses were matched with *in silico* tryptic digests of

Table 3 Number of proteins and peptides, as identified by mass spectrometry, matching the predicted proteomes of three different versions of the *Botrytis cinerea* genome, the pool of all three predicted proteomes and the six-frame translation of the gapless assembly.

	Protein IDs	Peptide IDs
v1 gene models	3405	17 057
v2 gene models	3195	16 664
v3 gene models	3560	19 363
v1 + v2 + v3 combined	3629	19 594
v3 six-frame translation	2239	14 769

entire proteomes predicted on the earlier versions of the B05.10 genome (v1 and v2) and the proteome predicted on the current, gapless genome (v3), as well as on a six-frame translation of the v3 genome assembly (Table 3). The number of proteins and peptides that were identified was substantially higher for the v3 proteome when compared with the earlier versions (Fig. 6). About 504 peptides could be detected in v3, which were not identified in v1 or v2. The 54 peptides that were uniquely identified in v1, but not in v2 and v3, were manually checked in order to verify whether any gene models present in v1 were overlooked in v3. The majority of these peptides were, in fact, present in proteins predicted in v3, but remained undetected because of their low Mascot scores. Only 10 of the 54 peptides specifically matching the v1 proteome corresponded to proteins that were missing or truncated in the v3 gene prediction.

SNP analysis of evolutionary pressure on coding sequences

SNP analysis between strains B05.10 and 09Bc11 was carried out to search for gene functions that might be under positive selection. The dN/dS ratio was calculated using PAML and the 10% of genes with the highest ratios were considered as potentially under evolutionary selection. This gene set did not show a significant over-representation of gene ontology terms or over-representation of putatively secreted proteins.

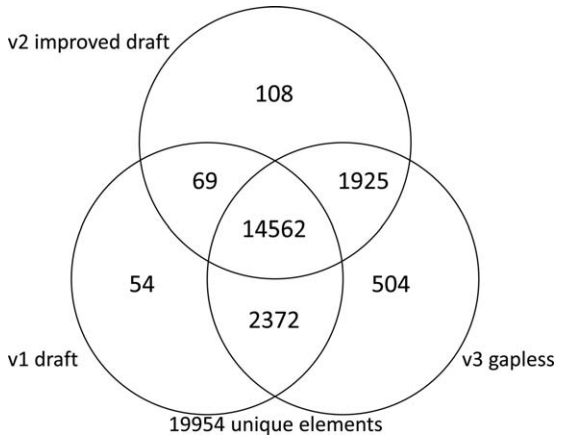

Fig. 6 Validation of gene predictions by proteome analysis. The numbers in the Venn diagram represent the numbers of peptides (derived from tryptic digests of soluble mycelium proteins) mapping to the three versions of the predicted proteome of *Botrytis cinerea* B05.10.

DISCUSSION

A combination of existing and new sequence data from Illumina and PacBio technology, supported by an optical map and a linkage map, was used to generate a near-finished, gapless assembly of the genome of *B. cinerea*, considered to be among the most important plant-pathogenic fungi in terms of economic and scientific relevance (Dean *et al.*, 2012). Isolate B05.10 is the most frequently used strain in molecular genetic studies on the virulence and developmental biology of this species (Antal *et al.*, 2012; Canessa *et al.*, 2013; Gonzalez *et al.*, 2013; Gronover *et al.*, 2005; van Kan *et al.*, 1997; Rui and Hahn, 2007; Schumacher *et al.*, 2014; Simon *et al.*, 2013; Stefanato *et al.*, 2009; Turrion-Gomez *et al.*, 2010; Viefhues *et al.*, 2014; Weiberg *et al.*, 2013).

Neither of the sequencing methods alone sufficed to generate an assembly of the quality that was obtained when combing PacBio with optical and linkage mapping. The PacBio assembly process resulted in the lowest number of contigs, but introduced several hundred indels, which caused inadvertent frameshifts in approximately 100 bona fide coding sequences (not shown). The assembly of Illumina sequence reads, even when performed with data from three different sized libraries (175 nucleotides, 3.5 and 10 kb), resulted in 100 gaps and approximately 1000 ambiguous base calls in low-repetitive sequences, and failed to include the majority of telomeric regions. Neither of the assembly methods alone could have included as many telomeric regions if they had not been supported by the optical map and linkage map. These observations concur with previous work on other filamentous pathogens with similar genome size and repeat content. *De novo* PacBio assemblies or assemblies that combine PacBio data with Illumina data, albeit often able to resolve chromosomes, are generally insufficient to generate gapless, near-complete genome assemblies without the use of optical mapping (Faino *et al.*, 2015; Seidl *et al.*, 2015). The genome assembly of *B. cinerea* presented here still contains nine unplaced contigs that represent telomeric regions which could not confidently be anchored, because of the lack of conspicuous, unique sequences that would allow them to be assigned to a specific chromosome. Therefore, the present assembly is gapless, but should be considered as near-finished.

The assembly comprises 18 contigs, 10 of which represent full chromosomes from telomere to telomere. The rDNA repeat is located at the beginning of Chr4. A single *B. cinerea* rDNA repeat unit is 9.0 kb in length and, from the size of sequences missing in the assembly (580 kb, as judged from the optical map), we

conclude that there are approximately 65 copies of the rDNA repeat in the *B. cinerea* genome. This is in line with the number of rDNA copies in ascomycete genomes, which has been reported to range from 45 to 150 (Ganley and Kobayashi, 2007). In addition to the rDNA repeats, eight telomeric regions that are missing in the assembly cover approximately 267 kb of sequence, which brings the entire genome size of *B. cinerea* B05.10 to 43.5 Mb. We propose the approximate position for the centromeric regions in 17 of the 18 chromosomes, but their exact locations and lengths remain to be confirmed by chromatin immunoprecipitation sequencing (CHIPseq), as carried out in other filamentous fungi (Smith *et al.*, 2012).

The genome contains two mini-chromosomes of 247 and 209 kb, respectively. Cytological studies by Shirane *et al.* (1989) on five *Botrytis* species indicated that four species analysed, including *B. cinerea*, contained 16 chromosomes, whereas the hybrid species *B. allii* contained 32 chromosomes. In addition, *Sclerotinia sclerotiorum*, the closest relative of *B. cinerea* for which a genome sequence is available, has been reported to possess 16 chromosomes (Amselem *et al.*, 2011). However, karyotype analysis by CHEF (contour-clamped homogeneous electrical field) gel electrophoresis of five *B. cinerea* isolates (not including B05.10) revealed the presence of one to three mini-chromosomes with sizes ranging from 0.22 to 0.58 Mb (van Kan *et al.*, 1993), similar to the small chromosomes observed in isolate B05.10. Such molecules are too small to be detected by standard microscopy as performed by Shirane *et al.* (1989).

Chr18 of strain B05.10 appeared to be lacking in strain 09Bc11; hence, half of the progeny from the cross between these two isolates did not contain any sequences derived from Chr18. Strikingly, however, the coverage of sequences derived from Chr17 and Chr18 in meiotic progeny was often lower when compared with the read coverage for 'core' chromosomes. Some progeny showed only approximately 25% of the read coverage for Chr17 and Chr18 when compared with other progeny, whereas the read coverage for Chr1–16 was even across all progeny. This can partly be explained by the high diversity between the parents for these chromosomes. The mapping of sequence reads from strain 09Bc11 onto B05.10 Chr17 showed a patchy pattern, with regions lacking mapped reads interspersed with regions showing properly mapped reads, suggesting a quite different architecture of the Chr17-derived sequences in strain 09Bc11. Progeny that inherit the 09Bc11 chromosome are therefore expected to have a lower read coverage. In addition, these mini-chromosomes may be mitotically unstable, for instance as a result of non-disjunction during mitosis resulting in lower coverage from the culture of genetically heterogeneous mycelium. Previous experiments identified non-Mendelian segregation of *B. cinerea* mini-chromosomes in meiosis. A clone derived from a 0.22-Mb mini-chromosome in strain SAS56 (van Kan *et al.*, 1993) was used as probe in dot blot hybridizations to examine mini-chromosome inheritance in four complete tetrads sampled from a cross between isolates either carrying (SAS56) or lacking (Bc29) this mini-chromosome. In two of the four tetrads analysed, DNA from all eight single ascospore progeny hybridized to the mini-chromosome-specific probe, whereas the other tetrads showed a 4 : 4 segregation of hybridization signal (T. W. Prins and J. A. L. van Kan, unpublished data), which could be indicative of non-disjunction for these chromosomes during meiosis.

The two mini-chromosomes contain only few predicted genes (18 and 14 genes, respectively), most of which have no similarity to known proteins and are poorly supported by RNAseq evidence (not shown). Three genes on Chr17 (Bcin17g00020–Bcin17g00040) have orthologues in *Sclerotinia sclerotiorum* Chr1 (SS1G09724–SS1G09722); the remainder of Chr17 has no homology to *S. sclerotiorum*. *Botrytis cinerea* Chr18 has no homology to *S. sclerotiorum* whatsoever. There are several examples of plant-pathogenic fungi possessing accessory chromosomes (e.g. *Fusarium oxysporum*, *Nectria haematococca*) which contain large sets of genes that are important for pathogenesis (Coleman *et al.*, 2009; Ma *et al.*, 2010). The relevance of the two *B. cinerea* mini-chromosomes in virulence is not obvious from their sequence or the nature of the genes that they harbour, especially when considering the complete absence of Chr18 from strain 09Bc11. The large diversity in sequence coverage for Chr17 and Chr18, and, especially, the very low coverage observed in some progeny, raises the question of how such seemingly unstable chromosomes are retained at all.

The genetic linkage map not only provided a robust tool to anchor, correct and validate the sequence assembly, but also yielded evidence for a translocation between strains B05.10 and 09Bc11, between Chr1 and Chr6. In addition, the linkage map provided insight into the correlation between the genetic and physical distances. The recombination rate in the cross varied between chromosomes, and ranged from 86 to 128 cM/Mb, which is on the same order of magnitude as in *Zymoseptoria tritici* (Croll *et al.*, 2015). Several recombination hotspots were detected: in some cases, the physical distance between markers that were 10 cM apart was only a few kilobases, whereas, in other cases, regions were identified that spanned many dozens of kilobases, yet did not show a single recombination event. The regions with high sexual recombination rates were not located close to telomeres, as reported for *Z. tritici* (Croll *et al.*, 2015), nor were they preferentially associated with the presence of transposable elements. With the full genome sequence of both parents available, it would be interesting to explore the underlying drivers of variations in recombination frequencies.

The quality of the present genome assembly and the predicted proteins was assessed by peptide identification of a soluble mycelial protein sample. The number of proteins identified in the v3

proteome was 3560, reflecting one-third of the total (predicted) gene products encoded in the *B. cinerea* genome in a single protein sample. It is likely that the analysis of protein extracts from additional tissue types (conidia, sclerotia, apothecia) or the use of different sampling methods (e.g. for membrane-bound proteins) would further increase the number of proteins identified. The number of detected peptides matching the predicted proteins in the three versions of the genome reflected the quality of the genomes and their gene models. The v1 draft genome with 10% Ns contained the highest number of predicted proteins (16 448), approximately 2800 of which were considered to be spurious (Amselem *et al.* 2011). Less than 50% of the 13 664 'high-confidence' v1 gene models were completely correct (J. A. L. van Kan, unpublished data). The improved draft assembly was more concordant (fewer scaffolds and only 1% Ns), but the number of peptides matching the predicted proteins of v2 was even lower than for v1, because the gene prediction for v2 was imprecise, leading to an underestimation of the predicted v2 proteome by about 10% (not shown). The present v3 proteome lacks only very few bona fide gene models that were identified in the soluble proteome analysis.

Manual inspection and curation of all gene models on Chr1 have already resulted in the correction of several incorrect gene predictions and revealed features for which the biological relevance remains to be explored: the occurrence of extremely unusual splice junctions that are clearly spliced efficiently; two regions in which gene models overlap over their entire length in opposite strands; the presence of spliced antisense RNAs overlapping coding sequences; and 5′-UTRs containing multiple introns, some with lengths beyond 1000 nucleotides. A community curation effort is ongoing to curate gene models on all chromosomes, culminating in further improvement of gene models and a genome-wide overview of the occurrence of such unusual gene structures, which may generate novel biological questions. Results of the community curation will be shared through the Ensembl-Fungi platform (http://fungi.ensembl.org/Botrytis_cinerea/) as they become available. The annotation of repetitive sequences and transposable elements (Amselem *et al.*, 2015) will result in a better understanding of the dynamics of these elements, and unveil their contribution to the production of small effector RNAs that are capable of modulating the plant immune response (Weiberg *et al.*, 2013).

The present genome assembly will serve as a valuable reference for comparison with other *B. cinerea* isolates, as well as related taxa. Several projects are ongoing to explore the diversity between *B. cinerea* and its close relatives (M. Hahn, University of Kaiserslautern, Germany, personal communication), more distant *Botrytis* species (J. A. L. van Kan, unpublished data) and members of the Sclerotiniaceae (J. Rollins, University of Florida, Gainesville, USA, and L. M. Kohn, University of Toronto, Canada, personal communication; R. Oliver, Curtin University, Australia, personal communication). Finally, this gapless reference genome can be considered as an essential milestone towards more systematic functional genomics in *B. cinerea*. This may become particularly relevant in the context of the fast development of genome editing techniques in multiple biological systems, including filamentous fungi (Arazoe *et al.*, 2015; Nødvig *et al.*, 2015).

EXPERIMENTAL PROCEDURES

Chromosomal DNA extraction and optical map data acquisition

Botrytis cinerea protoplasts were embedded in 0.5% low melting temperature agarose gel inserts (Schwartz and Cantor, 1984), and then lysed using 0.5 M ethylenediaminetetraacetic acid (EDTA), 1% lauroyl sarcosine, 2 mg/mL proteinase K, 1 M NaCl, pH 9.5, for two 24-h incubations at 50 °C. Then, the DNA gel inserts were washed overnight in TE [10 mM tris(hydroxymethyl)aminomethane (Tris), 1 mM EDTA, pH 8.0], melted at 72 °C for 7 min, followed by the addition of β-agarase (100 μL of TE + 1 Unit β-agarase; New England Biolabs, Ipswich, MA, USA), and incubated at 42 °C for 2 h. Dilutions were made with TE to ensure minimal crowding of molecules on optical mapping surfaces. Bacteriophage T3 DNA was added (10 pg/μL) as internal size standard. Samples were mounted onto an optical mapping surface and examined by fluorescence microscopy to check integrity and concentration. DNA molecules were mounted onto optical mapping surfaces using a silicone microchannel device (Dimalanta *et al.*, 2004; Zhou *et al.*, 2007) for the generation of single-molecule image datasets, which were automatically processed using a pipeline (Teague *et al.*, 2010) that constructed approximately 140 000 Rmaps (cleaved by *Bst*EII). The Rmap (single-molecule restriction maps) dataset was approximately 38.5 Gb, representing approximately 907× coverage of the *B. cinerea* genome.

Optical map *de novo* assembly and comparisons with sequence scaffolds

An optical map spanning the entire genome was assembled using the Optical Mapping System (Dimalanta *et al.*, 2004; Teague *et al.*, 2010; Zhou *et al.*, 2007). The previously described Rmap dataset was clustered using a k-mer hash to approximate a De Bruijn graph algorithm. Each of the clustered Rmaps was independently assembled using the optical map assembler (Valouev *et al.*, 2006a,b, c; Zhou *et al.*, 2007). The finished optical map (42.4 Mb) comprised 18 chromosomes. Each chromosomal (optical) contig was, on average, spanned at a depth of >80× by Rmaps, with the exception of Chr17 and Chr18 (Table 1).

Sexual crosses and isolation of single ascospore progeny

Sexual crosses were performed between strains 09Bc11 (maternal parent) and B05.10 (paternal parent) following the protocol of Faretra *et al.* (1988). Mature apothecia were sampled and crushed in water to release the ascospores. The spore suspension was filtered through glasswool to remove cell debris and plated at low density on agar plates. Single ascospore germlings

were sampled after 2 days and propagated for fungicide resistance assessment and DNA extraction. A total of 70 progeny was analysed.

Illumina sequencing of strain 09Bc11 and sexual progeny

One microgram of DNA was fragmented to approximately 400 bp using a Covaris S2 ultrasonicator (Covaris, Woburn, Massachusetts, USA). Illumina P5/P7 sequencing adapters were ligated using an Apollo 324 NGS Library Prep System with a PrepX ILM DNA Library Kit (WaferGen Biosystems, Fremont, California, USA). A unique multiplex index was added to each sample at the polymerase chain reaction (PCR) amplification step using KAPA HiFi HotStart ReadyMix (2×) (Kapa Biosystems Inc., Wilmington, Massachusetts, USA). Ten to twelve indexed samples were pooled and sequenced per lane on a HiSeq2000 (100 cycle paired-end runs) (Illumina Inc., San Diego, California, USA). The yield was 18 000 read pairs/sample, equivalent to 3.6 Gb/sample.

Illumina sequencing of strain B05.10, processing and assembly

Botrytis cinerea strain B05.10 DNA was isolated, sequenced and trimmed as described previously (Staats and van Kan, 2012), with the exception that the 3.5-kb insert library was trimmed as described below. In addition, a mate pair sequencing library with an estimated fragment size of 11.1 ± 1.4 kb was prepared and sequenced by Beijing Genomics Institute (BGI, Hong Kong) using Illumina HiSeq2000 technology (22 299 530 reads). Reads were trimmed to a length of 33 bp, starting at the fourth and ending at the 36th nucleotide, using fastx trimmer. Sequencing libraries were used as input for AllPathsLG (version 43019) (Gnerre *et al.*, 2011; Ribeiro *et al.*, 2012), which was run using default settings for a haploid genome. AllPathsLG produced 128 scaffolds of a total length of just over 42 Mb, with an N50 of 2436 kb (taking into account 1.3 million N characters).

The Illumina reads of the 70 sexual progeny were mapped on the v2 assembly of B05.10 (Staats and van Kan, 2012) to identify SNPs. Four progeny were removed from the analysis, three because they were identical to the B05.10 parent and one which was identical to one other progeny. SNPs were filtered if any of the following conditions were met: ambiguous score in any of the progeny; missing data in any of the progeny; less than 25 reference or less than 25 variant calls across all progeny; or the presence of ambiguous residues or sequence ends within 50 nucleotides of the SNP. After filtering, 75 117 high-confidence SNPs were retained as markers. These were mapped onto subsequent assemblies by aligning a short sequence of 101 nucleotides composed of 50 nucleotides on either side of the SNP and the B05.10 allele of the SNP itself using strict alignment criteria. Adjacent markers that showed the same pattern of inheritance over the sequenced progeny were merged into groups, yielding 2062 marker groups. These groups were used to calculate the most likely genetic map employing JoinMap (v 4.0). Regions with a genetic distance of >7.5 cM over a sequence of <20 kb, or a genetic distance of >15 cM over a sequence of <50 kb, were considered as recombination hotspots.

Comparison of the linkage map with the order of marker groups on the assembly identified incongruences in the assembly. Regions of incongruence were manually split, reassembled and validated by inspecting the mapping of paired-end reads over the merged region. Finally, gap closure was performed using a library of Sanger sequence reads employed for v1 of the B05.10 genome sequence (Amselem *et al.*, 2011) to fill the gaps. Briefly, sequences flanking either side of the gap were used to match Sanger reads. Matching reads were reassembled together with the flanking sequence into contigs using CAP3 (Huang and Madan, 1999). The resulting (extended) contigs were then used as input for subsequent iterative rounds of Sanger read matching and reassembly. The procedure removed 173 gaps (from a total of 803) larger than 20 nucleotides, removing 62 027 N characters, shortening the assembly by 17 149 bp.

PacBio sequencing of strain B05.10 and *de novo* assembly

High-molecular-weight DNA was isolated from protoplasts of strain B05.10 by gently lysing the protoplasts in 15 mL of 25 mM Tris-HCl, 10 mM EDTA (TE, pH 8.0) with 1% sodium dodecylsulfate (SDS) at 50 °C under gentle agitation. The lysate was centrifuged at 2000 g for 15 min and the pellet was discarded. One volume of isopropanol and one-tenth volume of 3 M sodium acetate (pH 5.5) were added to the supernatant and the precipitating DNA was removed with a glass hook. The precipitate was washed twice with 70% ethanol, dried, dissolved in 10 mL of TE and treated with 5 μg/mL RNAse A at 37 °C for 1 h. The sample was then extracted with buffer-saturated phenol under gentle agitation for 3 days, and centrifuged. The aqueous phase was extracted with chloroform under gentle agitation for 1 day. After centrifugation, the aqueous phase was recovered and DNA was precipitated by the addition of one volume of isopropanol and one-tenth volume of 3 M sodium acetate (pH 5.5). Precipitating DNA was retrieved with a glass hook, washed twice with 70% ethanol, dried and dissolved in 0.5 mL of 10 mM Tris-HCl, 0.1 mM EDTA (pH 7.5). Gel electrophoresis showed that the DNA obtained was larger than undigested phage λ DNA (48 kb). The yield of DNA was approximately 25 μg. DNA was concentrated using Ampure PB beads (Pacific Biosciences of California, Inc., Menlo Park, CA, USA). PacBio SMRTbell libraries were constructed and sequenced by Keygene NV (Wageningen, the Netherlands) according to the manufacturer's instructions (Pacific Biosciences). A total of 273 637 reads was obtained with a total number of read bases of 2.76×10^9 (post-filtering), of which the mean read length was 10 088 nucleotides. Sequence reads were assembled using HGAP (v3, Chin *et al.*, 2013) and the contigs were scaffolded by PBJelly (version 14.9.9, English *et al.*, 2012). After scaffolding, consensus was generated using Quiver (Chin *et al.*, 2013).

mRNA sequencing

mRNA from germlings, spores, mycelia, fungicide-treated and dimethylsulfoxide (DMSO)-treated mycelium was isolated from 3 μg of total RNA using the NEBNext® Poly(A) mRNA Magnetic Isolation Module (New England Biolabs, Ipswich, Massachusetts, USA). The resulting mRNA was fragmented to approximately 200 nucleotides using the NEBNext® Magnesium RNA Fragmentation Module and converted to cDNA using the NEBNext® RNA First Strand Synthesis Module plus the NEBNext® Second Strand Synthesis Module (New England Biolabs). The addition of Illumina sequencing adapters and multiplexing indices was performed using the method described above. Libraries were multiplexed and run in a single HiSeq lane in a 100-cycle paired-end run. The average number of read pairs per sample was 24 000 000 000.

RNA from sclerotia, apothecia (in different stages of development) and ascospores was isolated as described by Terhem (2015) and sequenced by BGI.

RNA was isolated from *B. cinerea* (isolate B05.10)-inoculated leaves and stem segments of tomato cultivar Moneymaker, its wild relative *Solanum habrochaites* (genotype Lyc4) and two introgression lines derived from a cross between Moneymaker and Lyc4 (Finkers *et al.*, 2007). Plant tissues were sampled for RNA extraction at 1 and 3 days post-inoculation. Sequencing was carried out at the Wageningen UR sequencing facility.

Gene prediction

Repeat masking was performed as described by Staats and van Kan (2012). For gene predictions, several gene callers were run and later integrated using EVM (Haas *et al.*, 2008). Genes were predicted using Genemark-HMM-ES (Ter-Hovhannisyan *et al.*, 2008), Augustus (Stanke *et al.*, 2006) and SNAP (Korf, 2004). Genemark-HMM-ES was self-trained on the full genome sequence or a version in which large repeats were masked. A training set based on approximately 1000 manually curated gene models from v2 (Staats and van Kan, 2012) was used to train Augustus and SNAP. Augustus was also run with intron hints, using GSNAP alignment of RNAseq reads of five distinct RNA samples.

RNA sequences were employed in genome-guided assembly using Trinity (Grabherr *et al.*, 2011) and PASA (Haas *et al.*, 2003), the output of which was also used as input for EVM. Finally, proteomes of 12 related fungal species (Table S5, see Supporting Information) were aligned to the assembly using the National Center for Biotechnology Information (NCBI) BLAST (Altschul *et al.*, 1990) for initial placement, followed by alignment using Exonerate to the region containing the best BLAST match and presented to EVM as separate tracks. In EVM, *ab initio* predictions were weighted as follows: SNAP 1, Naive Augustus 1, Augustus with hints 3 and Genemark on a masked or unmasked genome as 2 each. All Exonerate protein alignments were weighted as 1 per species. Alignments of PASA-assembled transcripts were weighted as 20. Gene models that had large overlap with repeat regions were removed. Gene models were further refined by manual curation. In cases in which splice variants were annotated, the isoform estimated to be most common based on RNAseq coverage levels was selected as representative.

Analysis for the presence of core eukaryotic genes was performed using the CEGMA pipeline, version 2.5 (Parra *et al.*, 2007).

In-solution digestion of proteins

Mycelium total protein concentration was measured using the Bradford protein assay [bovine serum albumin as standard; Bio-Rad Ltd.; Hemel Hempstead, Hertfordshire, United Kingdom]. A total of 500 μg of protein was transferred into a reaction tube and denatured in 8 M urea reduced with 10 mM dithiothreitol (DTT) for 45 min at 50 °C. Samples were alkylated in 50 mM iodoacetamide for 1 h in the dark at room temperature. Six volumes of pre-chilled acetone (−20 °C) were added to each sample to precipitate protein overnight, prior to trypsin digestion. Acetone was decanted and samples were left to air dry for 20 min. Pellets were resuspended in 20 μL of tetrafluoroethylene, followed by 20 μL of 50 mM tetraethylammonium bicarbonate (TEAB), and trypsin was added at an enzyme to protein ratio of 1 : 100 (w/w). Incubation took place overnight at 37 °C with agitation. OFFGEL isoelectric focusing in-solution peptide separation was carried out using the 24-well set-up as instructed in the manufacturer's instruction manual (Agilent Technologies LDA UK Ltd.; Stockport, Cheshire, United Kingdom). Following separation, Thermo Scientific Pierce (Fisher Scientific – UK Ltd.; Bishop Meadow Road, Loughborough, United Kingdom) C18 Tips (100 μL) were used to enable fast and efficient peptide capture, concentration, desalting and elution. All OFFGEL fractions were then dried down and stored at −20 °C.

MS analysis

Samples were run on a Q-Exactive Plus mass spectrometer coupled to a Dionex Ultimate 3000 RSLC nano system (Thermo Fisher Scientific Ltd.; Hemel Hempstead, Hertfordshire, United Kingdom). Prior to injection, samples were resuspended in 50 μL of 0.1% formic acid. An injection of 5 μL was loaded onto a C18 PepMap 300 Å trap column, 200 μm × 5 mm (5 μm bead size). Reversed-phase chromatographic separation was performed on a 75-μm i.d. × 50-cm column (2 μm bead size) using a 90-min linear gradient of 3%–40% solvent B (CH_3CN 100% + 0.1% formic acid) with a flow rate of 300 nL/min. The mass spectrometer was operated in a data-dependent mode to automatically switch between Orbitrap MS and MS/MS acquisition. Survey full-scan spectra (from *m/z* 300 to 1800) were acquired in the Orbitrap with a resolution of 70 000 at *m/z* 200 and an Fourier transform target value of 1×10^6 ions. The 10 most intense ions were sequentially selected for fragmentation using higher energy collisional dissociation (HCD) and dynamically excluded for 30 s. Fragmented ions were scanned in the Orbitrap at a resolution of 17 500 at *m/z* 200. For accurate mass measurement, the lock mass option was enabled using the polydimethylcyclosiloxane ion (*m/z* 445.120025) as an internal calibrant. Data processing was carried out using Proteome Discoverer 1.4 (Thermo Scientific), and database searching was conducted on an in-house installation of Mascot Server 2.5 (Matrix Science Ltd.; London, United Kingdom). Mascot settings were as follows: precursor mass tolerance, 3 ppm; fragment mass tolerance, 0.5 Da; maximum missed cleavage sites, 2; ion score or expected cut-off, 20; significance threshold, $P < 0.01$; cysteine carbamidomethylation (CAM-C) was chosen as fixed post-translational modification (PTM), whereas methionine oxidation (Met-Ox) was chosen as variable PTM.

Data storage

The genome is available at the Ensembl-Fungi platform (http://fungi.ensembl.org/Botrytis_cinerea/); the underlying data can be downloaded from the European Bioinformatics Institute (EBI). Nucleotide sequences of Chr1–Chr18 are deposited in GenBank under accession numbers CP009805–CP009822, BioProject number PRJNA264284.

ACKNOWLEDGEMENTS

The authors declare that they have no conflicts of interest. MFS acknowledges the receipt of a VENI grant from the Research Council Earth and Life Sciences (ALW) of the Netherlands Organization of Scientific Research (NWO), project number 863.15.005. The optical map work by SZ and DCS was supported by grants from NIF (USA) and Syngenta. PacBio sequencing of strain B05.10 was performed by Alexander Wittenberg and Peter de Heer (Keygene, Wageningen, the Netherlands). Helder Pedro, Uma Maheswari and Paul Kersey of the European Bioinformatics

Institute (EBI) are gratefully acknowledged for their generous assistance in presenting data in Ensembl-Fungi and facilitating the community curation platform. Colleagues in the *Botrytis* community are acknowledged for fruitful discussions and for their participation in the community annotation.

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.

Amselem, J., Cuomo, C.A., van Kan, J.A.L., Viaud, M., Benito, E.P., Couloux, A., Coutinho, P.M., de Vries, R.P., Dyer, P.S., Fillinger, S., Fournier, E., Gout, L., Hahn, M., Kohn, L., Lapalu, N., Plummer, K.M., Pradier, J.-M., Quévillon, E., Sharon, A., Simon, A., ten Have, A., Tudzynski, B., Tudzynski, P., Wincker, P., Andrew, M., Anthouard, V., Beever, R.E., Beffa, R., Benoit, I., Bouzid, O., Brault, B., Chen, Z., Choquer, M., Collémare, J., Cotton, P., Danchin, E.G., Da Silva, C., Gautier, A., Giraud, C., Giraud, T., Gonzalez, C., Grossetete, S., Güldener, U., Henrissat, B., Howlett, B.J., Kodira, C., Kretschmer, M., Lappartient, A., Leroch, M., Levis, C., Mauceli, E., Neuvéglise, C., Oeser, B., Pearson, M., Poulain, J., Poussereau, N., Quesneville, H., Rascle, C., Schumacher, J., Ségurens, B., Sexton, A., Silva, E., Sirven, C., Soanes, D.M., Talbot, N.J., Templeton, M., Yandava, C., Yarden, O., Zeng, Q., Rollins, J.A., Lebrun, M.-H. and Dickman, M. (2011) Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. *PLoS Genet.* **7**, e1002230.

Amselem, J., Lebrun M.-H. and Quesneville, H. (2015) Whole genome comparative analysis of transposable elements provides new insight into mechanisms of their inactivation in fungal genomes. *BMC Genomics*, **16**, 141.

Antal, Z., Rascle, C., Cimerman, A., Viaud, M., Billon-Grand, G., Choquer, M. and Bruel, C. (2012) The Homeobox BcHOX8 gene in *Botrytis cinerea* regulates vegetative growth and morphology. *PLoS One*, **7**, e48134.

Arazoe, T., Miyoshi, K., Yamato, T., Ogawa, T., Ohsato, S., Arie, T. and Kuwata, S. (2015) Tailor-made CRISPR/Cas system for highly efficient targeted gene replacement in the rice blast fungus. *Biotechnol. Bioeng.* **112**, 2543–2549.

Canessa, P., Schumacher, J., Hevia, M.A., Tudzynski, P. and Larrondo, L.F. (2013) Assessing the effects of light on differentiation and virulence of the plant pathogen *Botrytis cinerea*: characterization of the White Collar complex. *PLoS One*, **8**, e84223.

Chin, C.S., Alexander, D.H., Marks, P., Klammer, A.A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E.E., Turner, S.W. and Korlach, J. (2013) Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods*, **10**, 563–569.

Coleman, J.J., Rounsley, S.D., Rodriguez-Carres, M., Kuo, A., Wasmann, C.C., Grimwood, J., Schmutz, J., Taga, M., White, G.J., Zhou, S., Schwartz, D.C., Freitag, M., Ma, L.-J., Danchin, E.G.J., Henrissat, B., Coutinho, P.M., Nelson, D.R., Straney, D., Napoli, C.A., Barker, B.M., Gribskov, M., Rep, M., Kroken, S., Molnár, I., Rensing, C., Kennell, J.C., Zamora, J., Farman, M.L., Selker, E.U., Salamov, A., Shapiro, H., Pangilinan, J., Lindquist, E., Lamers, C., Grigoriev, I.V., Geiser, D.M., Covert, S.F., Temporini, E. and VanEtten, H.D. (2009) The genome of *Nectria haematococca*: contribution of supernumerary chromosomes to gene expansion. *PLoS Genet.*, **5**, e1000618.

Croll, D., Lendenmann, M.H., Stewart, E. and McDonald, B.A. (2015) The impact of recombination hotspots on genome evolution of a fungal plant pathogen. *Genetics*, **201**, 1213–1228.

Dalmais, B., Schumacher, J., Moraga, J., Le Pecheur, P., Tudzynski, B., Collado, I.G. and Viaud, M. (2011) The *Botrytis cinerea* phytotoxin botcinic acid requires two polyketide synthases for production and has a redundant role in virulence with botrydial. *Mol. Plant Pathol.* **12**, 564–579.

Dean, R., van Kan, J.A.L., Pretorius, Z.A., Hammond-Kosack, K.E., Di Pietro, A., Spanu, P.D., Rudd, J.J., Dickman, M., Kahmann, R. and Ellis, J. (2012) The Top 10 fungal pathogens in molecular plant pathology. *Mol. Plant Pathol.* **13**, 414–430.

Dimalanta, E.T., Lim, A., Runnheim, R., Lamers, C., Churas, C., Forrest, D.K., de Pablo, J.J., Graham, M.D., Coppersmith, S.N., Goldstein, S. and Schwartz, D.C. (2004) A microfluidic system for large DNA molecule arrays. *Anal. Chem.* **76**, 5293–5301.

Elad, Y., Pertot, I., Cotes Prado, A.M. and Stewart, A. (2015) Plant hosts of *Botrytis* spp. In: *Botrytis – The Fungus, the Pathogen and its Management in Agricultural Systems* (**Fillinger, S.** and **Elad, Y.**, eds.), pp. 413–486. Berlin: Springer.

English, A.C., Richards, S., Han, Y., Wang, M., Vee, V., Qu, J., Qin, X., Muzny, D.M., Reid, J.G., Worley, K.C. and Gibbs, R.A. (2012) Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PLoS One*, **7**, e47768.

Faino, L., Seidl, M.F., Datema, E., van den Berg, G.C.M., Janssen, A., Wittenberg, A.H.J. and Thomma, B.P.H.J. (2015) Single-Molecule Real-Time sequencing combined with optical mapping yields completely finished fungal genome. *mBio*, **6**, e00936–15.

Faretra, F., Antonacci, E. and Pollastro, S. (1988) Sexual behavior and mating system of *Botryotinia fuckeliana*, teleomorph of *Botrytis cinerea*. *J. Gen. Microbiol.* **134**, 2543–2550.

Finkers, R., van Heusden, A.W., Meijer-Dekens, F., van Kan, J.A.L., Maris, P. and Lindhout, P. (2007) The construction of a *Solanum habrochaites* LYC4 introgression line population and the identification of QTLs for resistance to *Botrytis cinerea*. *Theor. Appl. Genet.* **114**, 1071–1080.

Ganley, A.R.D. and Kobayashi, T. (2007) Highly efficient concerted evolution in the ribosomal DNA repeats: total rDNA repeat variation revealed by whole-genome shotgun sequence data. *Genome Res.* **17**, 184–191.

Gnerre, S., MacCallum, I., Przybylski, D., Ribeiro, F.J., Burton, J.N., Walker, B.J., Sharpe, T., Hall, G., Shea, T.P., Sykes, S., Berlin, A.M., Aird, D., Costello, M., Daza, R., Williams, L., Nicol, R., Gnirke, A., Nusbaum, C., Lander, E.S. and Jaffe, D.B. (2011) High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc. Natl. Acad. Sci. USA*, **108**, 1513–1518.

Gonzalez, M., Brito, N., Frias, M. and Gonzalez, C. (2013) *Botrytis cinerea* protein O-mannosyltransferases play critical roles in morphogenesis, growth, and virulence. *PLoS One*, **8**, e65924.

Grabherr, M.G., Haas, B.J, Yassour, M., Levin, J.Z., Thompson, D.A, Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W, Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652.

Gronover, C.S., Schumacher, J., Hantsch, P. and Tudzynski, B. (2005) A novel seven-helix transmembrane protein BTP1 of *Botrytis cinerea* controls the expression of GST-encoding genes, but is not essential for pathogenicity. *Mol. Plant Pathol.* **6**, 243–256.

Haas, B.J., Delcher, A.l., Mount, S.M., Wortman, J.R., Smith, R.K., Hannick, L.I., Maiti, R., Ronning, C.M., Rusch, D.B., Town, C.D., Salzberg, S.L., White, O. (2003) Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* **31**, 5654–5666.

Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C.R., Wortman, J.R. (2008) Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**, R7.

Hahn, M. (2014) The rising threat of fungicide resistance in plant pathogenic fungi: *Botrytis* as a case study. *J. Chem. Biol.* **7**, 133–141.

Huang, X.Q. and Madan, A. (1999) CAP3: a DNA sequence assembly program. *Genome Res.* **9**, 868–877.

van Kan, J.A.L. (2006) Licensed to kill: the lifestyle of a necrotrophic plant pathogen. *Trends Plant Sci.* **11**, 247–253.

van Kan, J.A.L., Goverse, A. and van der Vlugt-Bergmans, C.J.B. (1993) Electrophoretic karyotype analysis of *Botrytis cinerea*. *Neth. J. Plant Pathol.* **99**, 119–128.

van Kan, J.A.L., van't Klooster, J.W., Dees, D.C.T., Wagemakers, C.A.M. and van der Vlugt-Bergmans, C.J.B. (1997) Cutinase A of *Botrytis cinerea* is expressed, but not essential, during penetration of gerbera and tomato. *Mol. Plant–Microbe Interact.* **10**, 30–38.

van Kan, J.A.L., Shaw, M.A. and Grant-Downton, R.T. (2014) *Botrytis* species: relentless necrotrophic thugs or endophytes gone rogue? *Mol. Plant Pathol.* **15**, 957–961.

Korf, I. (2004) Gene finding in novel genomes. *BMC Bioinformatics*, **5**, 59.

Leroch, M., Plesken, C., Weber, R.W.S., Kauff, F., Scalliet, G. and Hahn, M. (2013) Gray mold populations in German strawberry fields are resistant to multiple fungicides and dominated by a novel clade closely related to *Botrytis cinerea*. *Appl. Environ. Microbiol.* **79**, 159–167.

Leroux, P., Fritz, R., Debieu, D., Albertini, C., Lanen, C., Bach, J., Gredt, M. and Chapeland, F. (2002) Mechanisms of resistance to fungicides in field strains of *Botrytis cinerea*. *Pest Manag. Sci.* **58**, 876–888.

Leroux, P., Gredt, M., Leroch, M. and Walker, A.S. (2010) Exploring mechanisms of resistance to respiratory inhibitors in field strains of *Botrytis cinerea*, the causal agent of gray mold. *Appl. Environ. Microbiol.* **76**, 6615–6630.

Ma, L.J., van der Does, H.C., Borkovich, K.A., Coleman, J.J., Daboussi, M.J., Di Pietro, A., Dufresne, M., Freitag, M., Grabherr, M., Henrissat, B., Houterman, P.M., Kang, S., Shim, W.B., Woloshuk, C., Xie, X., Xu, J.R., Antoniw, J., Baker, S.E., Bluhm, B.H., Breakspear, A., Brown, D.W., Butchko, R.A., Chapman, S., Coulson, R., Coutinho, P.M., Danchin, E.G., Diener, A.,

Gale, L.R., Gardiner, D.M., Goff, S., Hammond-Kosack, K.E., Hilburn, K., Hua-Van, A., Jonkers, W., Kazan, K., Kodira, C.D., Koehrsen, M., Kumar, L., Lee, Y.H., Li, L., Manners, J.M., Miranda-Saavedra, D., Mukherjee, M., Park, G., Park, J., Park, S.Y., Proctor, R.H., Regev, A., Ruiz-Roldan, M.C., Sain, D., Sakthikumar, S., Sykes, S., Schwartz, D.C., Turgeon, B.G., Wapinski, I., Yoder, O., Young, S., Zeng, Q., Zhou, S., Galagan, J., Cuomo, C.A., Kistler, H.C. and Rep, M. (2010) Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature*, **464**, 367–373.

Nødvig, C.S., Nielsen, J.B., Kogle, M.E. and Mortensen, U.H. (2015) A CRISPR-Cas9 system for genetic engineering of filamentous fungi. *PLoS One*, **10**, e0133085.

Pinedo, C., Wang, C.M., Pradier, J.-M., Dalmais, B., Choquer, M., Le Pecheur, P., Morgant, G., Collado, I.G., Cane, D.E. and Viaud, M. (2008) Sesquiterpene synthase from the botrydial biosynthetic gene cluster of the phytopathogen *Botrytis cinerea*. *ACS Chem. Biol.* **3**, 791–801.

Parra, G., Bradnam, K. and Korf, I. (2007) CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics*, **23**:1061–1067.

Ribeiro, F.J., Przybylski, D., Yin, S.Y., Sharpe, T., Gnerre, S., Abouelleil, A., Berlin, A.M., Montmayeur, A., Shea, T.P., Walker, B.J., Young, S.K., Russ, C., Nusbaum, C., MacCallum, I. and Jaffe, D.B. (2012) Finished bacterial genomes from shotgun sequence data. *Genome Res.* **22**, 2270–2277.

Rui, O. and Hahn, M. (2007) The Slt2-type MAP kinase Bmp3 of *Botrytis cinerea* is required for normal saprotrophic growth, conidiation, plant surface sensing and host tissue colonization. *Mol. Plant Pathol.* **8**, 173–184.

Schumacher, J., Simon, A., Cohrs, K.C., Viaud, M. and Tudzynski, P. (2014) The transcription factor BcLTF1 regulates virulence and light responses in the necrotrophic plant pathogen *Botrytis cinerea*. *PLoS Genet.* **10**, e1004040.

Schwartz, D.C. and Cantor, C.R. (1984) Separation of yeast chromosome-sized DNAs by pulsed-field gradient gel-electrophoresis. *Cell*, **37**, 67–75.

Seidl, M.F., Faino, L., Shi-Kunne, X., van den Berg, G.C.M., Bolton, M.D. and Thomma, B.P.H.J. (2015) The genome of the saprophytic fungus *Verticillium tricorpus* reveals a complex effector repertoire resembling that of its pathogenic relatives. *Mol. Plant–Microbe Interact.* **28**, 362–373.

Shirane, N., Masuka, M. and Hayashi, Y. (1989) Light microscopic observation of nuclei and mitotic chromosomes of *Botrytis* species. *Phytopathology*, **79**, 728–730.

Simon, A., Dalmais, B., Morgant, G. and Viaud, M. (2013) Screening of a *Botrytis cinerea* one-hybrid library reveals a Cys_2His_2 transcription factor involved in the regulation of secondary metabolism gene clusters. *Fung. Genet. Biol.* **52**, 9–19.

Smith, K.M., Galazka, J.M., Phatale, P.A., Connolly, L.R. and Freitag, M. (2012) Centromeres of filamentous fungi. *Chromosome Res.* **20**, 635–656.

Sowley, E.N.K., Dewey, F.M. and Shaw, M.W. (2009) Persistent, symptomless, systemic, and seed-borne infection of lettuce by *Botrytis cinerea*. *Eur. J. Plant Pathol.* **126**, 61–71.

Spatafora, J., Stajich, J. and Grigoriev, I. (2013) 1000 Fungal Genomes project. *Phytopathology*, **103**, 137.

Staats, M. and van Kan, J.A.L. (2012) Genome update of *Botrytis cinerea* strains B05.10 and T4. *Eukaryotic Cell*, **11**, 1413–1414.

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S. and Morgenstern, B. (2006) AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**, W435–W439.

Stefanato, F.L., Abou-Mansour, E., Buchala, A., Kretschmer, M., Mosbach, A., Hahn, M., Bochet, C.G., Metraux, J.P. and Schoonbeek, H.-J. (2009) The ABC transporter BcatrB from *Botrytis cinerea* exports camalexin and is a virulence factor on *Arabidopsis thaliana*. *Plant J.* **58**, 499–510.

Teague, B., Waterman, M.S., Goldstein, S., Potamousis, K., Zhou, S., Reslewic, S., Sarkar, D., Valouev, A., Churas, C., Kidd, J.M., Kohn, S., Runnhein, R., Lamers, C., Forrest, D., Newton, M.A., Eichler, E.E., Kent-First, M., Surti, U., Livny, M., and Schwartz, D.C. (2010) High-resolution human genome structure by single molecule analysis. *Proc. Natl. Acad. Sci. USA*, **107**, 10 848–10 853.

Terhem, R.B. (2015) Sexual development of *Botrytis* species. PhD Thesis, Wageningen University.

Ter-Hovhannisyan, V., Lomsadze, A., Chernoff, Y.O. and Borodovsky, M. (2008) Gene prediction in novel fungal genomes using an *ab initio* algorithm with unsupervised training. *Genome Res.* **18**, 1979–1990.

Thomma, B.P.H.J., Seidl, M.F., Shi-Kunne, X., Cook, D., Bolton, M.D., van Kan, J.A.L. and Faino, L. (2016) Mind the gap; seven reasons to close fragmented genome assemblies. *Fung. Genet. Biol.* **90**, 24–30.

Turrion-Gomez, J.L., Eslava, A.P. and Benito, E.P. (2010) The flavohemoglobin BCFHG1 is the main NO detoxification system and confers protection against nitrosative conditions but is not a virulence factor in the fungal necrotroph *Botrytis cinerea*. *Fung. Genet. Biol.* **47**, 484–496.

Valouev, A., Schwartz, D., Zhou, S. and Waterman, M.S. (2006a) An algorithm for assembly of ordered restriction maps from single DNA molecules. *Proc. Natl. Acad. Sci. USA*, **103**, 15 770–15 775.

Valouev, A., Zhang, Y., Schwartz, D.C. and Waterman, M.S. (2006b) Refinement of optical map assemblies. *Bioinformatics*, **22**, 1217–1224.

Valouev, A., Li, L., Liu, Y., Schwartz, D.C., Yang, Y., Zhang, Y. and Waterman, M.S. (2006c) Alignment of optical maps. *J. Comput. Biol.* **13**, 442–446.

Viefhues, A., Heller, J., Temme, N. and Tudzynski, P. (2014) Redox systems in *Botrytis cinerea*: impact on development and virulence. *Mol. Plant–Microbe Interact.* **27**, 858–874.

Wang, Z. and Sachs, M.S. (1996) Ribosome stalling is responsible for arginine-specific translational attenuation in *Neurospora crassa*. *Mol. Cell. Biol.* **17**, 4904–4913.

Wang, Z., Fang, P. and Sachs, M.S. (1998) The evolutionarily conserved eukaryotic arginine attenuator peptide regulates the movement of ribosomes that have translated it. *Mol. Cell. Biol.* **18**, 7528–7536.

Wang, Z., Gaba, A. and Sachs, M.S. (1999) A highly conserved mechanism of regulated ribosome stalling mediated by fungal arginine attenuator peptides that appears independent of the charging status of arginyl-tRNAs. *J. Biol. Chem.* **274**, 37 565–37 574.

Weiberg, A., Wang, M., Lin, F.M., Zhao, H.W., Zhang, Z.H., Kaloshian, I., Huang H.D. and Jin, H.L. (2013) Fungal small RNAs suppress plant immunity by hijacking host RNA interference pathways. *Science*, **342**, 118–123.

Williamson, B., Tudzynski, B., Tudzynski, P. and van Kan, J.A.L. (2007) *Botrytis cinerea*: the cause of grey mould disease. *Mol. Plant Pathol.* **8**, 561–580.

Zhou, S., Herschleb, J., and Schwartz, D.C. (2007) A single molecule system for whole genome analysis. In: *New High Throughput Technologies for DNA Sequencing and Genomics*, Vol. 2 (Mitchelson, K.R., ed), pp. 265–300. Amsterdam: Elsevier Scientific Publishers.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Fig. S1 Alignments between the optical map (orange, top) and an *in silico* digest of the sequence assembly (grey, bottom) for all chromosomes. Each vertical line reflects the presence of a *Bst*EII restriction site.

Fig. S2 Correlation between physical and genetic distances for all chromosomes. The *x*-axis displays the genetic distance in centimorgan (cM) and the *y*-axis displays the physical distance in base pairs (bp). The correlation between physical and genetic distances and the correlation coefficients for separate chromosomes are given in the top right-hand corners of each graph. Each blue diamond represents a marker group [set of single nucleotide polymorphisms (SNPs) fully co-segregating in all progeny]. The green hatched arrow indicates the approximate position of a centromere [for chromosome 1 (Chr1), two possible positions of the centromere are indicated]. Black arrows indicate regions with high recombination rates.

Table S1 Proposed locations of centromeric regions.

Table S2 Regions in the *Botrytis cinerea* genome with high sexual recombination rates.

Table S3 Genes on chromosome 1 (Chr1) with an untranslated region (UTR) length of >1000 nucleotides and/or at least two introns in the UTR.

Table S4 Examples of spliced antisense RNAs that occur in the coding sequence.

Table S5 Information on 12 fungal proteomes used in gene prediction and annotation.