

# Genome Sequencing and Mapping Reveal Loss of Heterozygosity as a Mechanism for Rapid Adaptation in the Vegetable Pathogen *Phytophthora capsici*

Kurt H. Lamour,<sup>1</sup> Joann Mudge,<sup>2</sup> Daniel Gobena,<sup>1</sup> Oscar P. Hurtado-Gonzales,<sup>3</sup> Jeremy Schmutz,<sup>4,5</sup> Alan Kuo,<sup>4</sup> Neil A. Miller,<sup>6</sup> Brandon J. Rice,<sup>2</sup> Sylvain Raffaele,<sup>7</sup> Liliana M. Cano,<sup>7</sup> Arvind K. Bharti,<sup>2</sup> Ryan S. Donahoo,<sup>8</sup> Sabra Finley,<sup>1</sup> Edgar Huitema,<sup>9,10</sup> Jon Hulvey,<sup>11</sup> Darren Platt,<sup>4</sup> Asaf Salamov,<sup>4</sup> Alon Savidor,<sup>12</sup> Rahul Sharma,<sup>13,14,15</sup> Remco Stam,<sup>9,10</sup> Dylan Storey,<sup>1</sup> Marco Thines,<sup>13,14,15</sup> Joe Win,<sup>7</sup> Brian J. Haas,<sup>16</sup> Darrell L. Dinwiddie,<sup>6,17</sup> Jerry Jenkins,<sup>4,5</sup> James R. Knight,<sup>18</sup> Jason P. Affourtit,<sup>18</sup> Cliff S. Han,<sup>19</sup> Olga Chertkov,<sup>19</sup> Erika A. Lindquist,<sup>4</sup> Chris Detter,<sup>19</sup> Igor V. Grigoriev,<sup>5</sup> Sophien Kamoun,<sup>7</sup> and Stephen F. Kingsmore<sup>6,17</sup>

<sup>1</sup>University of Tennessee, Knoxville 37996, U.S.A.; <sup>2</sup>National Center for Genome Resources, Santa Fe, NM 87505, U.S.A.; <sup>3</sup>Pioneer Hi-Bred International, Johnston, IA 50131, U.S.A.; <sup>4</sup>United States Department of Energy Joint Genome Institute, Walnut Creek, CA 94598, U.S.A.; <sup>5</sup>Hudson Alpha Institute of Biotechnology, Huntsville, AL 35806, U.S.A.; <sup>6</sup>Children's Mercy Hospital, Kansas City, MO 64108, U.S.A.; <sup>7</sup>The Sainsbury Laboratory, Norwich NR4 7UH, U.K.; <sup>8</sup>University of Florida, IFAS-SWFREC, Immokalee 34142, U.S.A.; <sup>9</sup>Division of Plant Science, University of Dundee, Invergowrie, Dundee DD2 5DA, U.K.; <sup>10</sup>Plant Pathology Program, James Hutton Institute, Invergowrie, Dundee DD2 5DA, U.K.; <sup>11</sup>University of Massachusetts-Amherst, Department of Plant, Soil, and Insect Sciences, Amherst 01003, U.S.A.; <sup>12</sup>Department of Molecular Biology and Ecology of Plants, Tel Aviv University, Tel Aviv 69978, Israel; <sup>13</sup>Biodiversity and Climate Research Centre, D-60325 Frankfurt am Main, Germany; <sup>14</sup>Senckenberg Gesellschaft für Naturforschung, D-60325 Frankfurt am Main, Germany; <sup>15</sup>Goethe University, Department of Biological Sciences, Institute of Ecology, Evolution and Diversity, D-60323, Frankfurt am Main, Germany; <sup>16</sup>Broad Institute, Cambridge, MA 02142, U.S.A.; <sup>17</sup>School of Medicine, University of Missouri-Kansas City, Kansas City, MO 64108, U.S.A.; <sup>18</sup>Roche Applied Science, Branford, Connecticut 06405, U.S.A.; <sup>19</sup>Los Alamos National Laboratory, Department of Energy Joint Genome Institute, Los Alamos, NM 87545, U.S.A.

Submitted 7 February 2012. Accepted 7 June 2012.

**The oomycete vegetable pathogen *Phytophthora capsici* has shown remarkable adaptation to fungicides and new hosts. Like other members of this destructive genus, *P. capsici* has an explosive epidemiology, rapidly producing massive numbers of asexual spores on infected hosts. In addition, *P. capsici* can remain dormant for years as sexually recombined oospores, making it difficult to produce crops at infested sites, and allowing outcrossing populations to maintain significant genetic variation. Genome sequencing, development of a high-density genetic map, and integrative genomic or genetic characterization of *P. capsici* field isolates and intercross progeny revealed significant mitotic loss of heterozygosity (LOH) in diverse isolates. LOH was detected in clonally propagated field isolates and sexual progeny, cumulatively affecting >30% of the genome. LOH altered genotypes for more than 11,000 single-nucleotide variant sites and showed a strong association with changes in mating type and pathogenicity. Overall, it appears that LOH may provide a rapid mechanism for fixing alleles and may be an important component of adaptability for *P. capsici*.**

*Phytophthora capsici* is a virulent, hemibiotrophic pathogen of vegetable crops, inflicting significant losses worldwide (Erwin and Ribeiro 1996; Gevens et al. 2008; Lamour and Kamoun 2009; Lamour et al. 2012; Leonian 1922; Meitz et al. 2010). Its main hosts are pepper and cucurbits. During the last 90 years, *P. capsici* has spread geographically as agricultural cultivation of these hosts has intensified, dynamically adapting to fungicides and new hosts (Erwin and Ribeiro 1996; Gevens et al. 2008; Gobena et al. 2012; Hausbeck and Lamour 2004; Hurtado-Gonzales et al. 2008; Lamour and Hausbeck 2001; Lamour and Kamoun 2009; Lamour et al. 2012; Leonian 1922; Meitz et al. 2010). Like other members of this destructive genus (e.g., the late blight pathogen *P. infestans* and the sudden oak death pathogen *P. ramorum*), *P. capsici* has an explosive epidemiology, rapidly producing massive numbers of asexual spores on infected hosts (Hausbeck and Lamour 2004; Lamour and Hausbeck 2001; Lamour and Kamoun 2009) (Fig. 1C to G). In addition, *P. capsici* often features meiosis between two mating types, producing thick-walled oospores that remain dormant for years, making it very difficult to produce crops at infested sites, and allowing outcrossing populations to maintain significant genetic variation (Fig. 1A and B). Members of this genus are notoriously plastic in field and laboratory scenarios and we examined the molecular basis for rapid adaptation in *P. capsici* through genome sequencing, development of a high-density genetic map, and integrative genomic and genetic characterization of *P. capsici* field isolates and intercross progeny. Development of these tools and subsequent analyses revealed significant mitotic loss of heterozygosity (LOH) in clonally propagated field isolates as well as

K. H. Lamour and J. Mudge contributed equally to this work.

This Whole Genome Shotgun project has been deposited in the DDBJ, EMBL, and GenBank databases under accession ADVJ00000000. The version described in this paper is the first version, ADVJ01000000.

Corresponding author: K. Lamour; E-mail: klamour@utk.edu

\*The e-Xtra logo stands for "electronic extra" and indicates that 17 supplementary tables and 11 supplementary figures are published online. Also, Figures 1 through 5 appear in color online.

multiple sexual progeny. Follow-up tests revealed a strong association of LOH with changes in mating type and loss of pathogenicity.

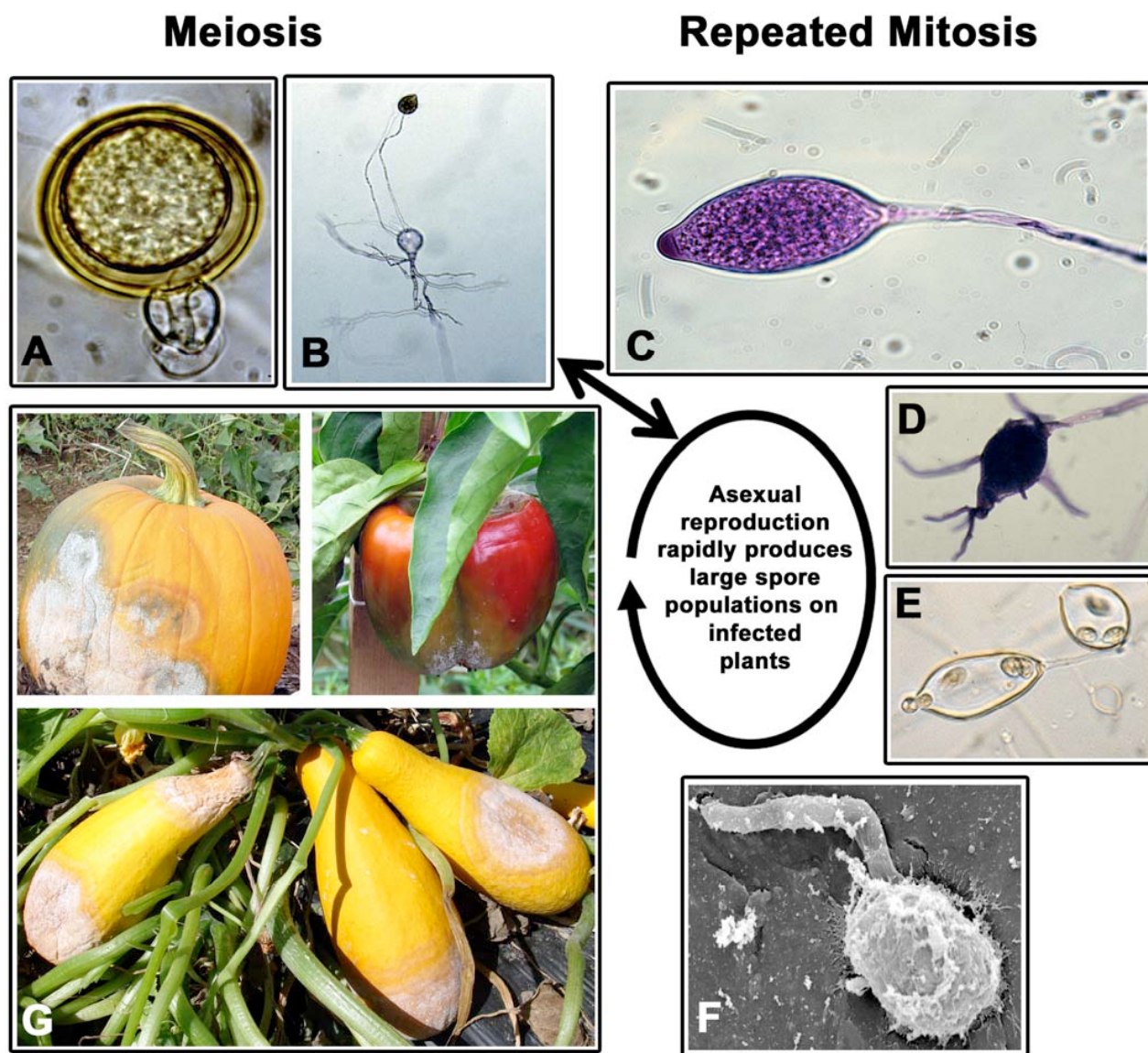
## RESULTS

### Genome sequencing, assembly, and annotation.

Assembling the nuclear genome of a diploid organism with frequent nucleotide variants can be difficult and a partially inbred line of *P. capsici* (LT1534) was developed for sequencing (Supplementary Fig. S1) (Gobena et al. 2012; Hurtado-Gonzales and Lamour 2009). The 64-Mbp reference genome was assembled using Arachne from 30× paired and singleton 454 FLX and Titanium genomic DNA reads, 5× Sanger paired reads, 56,448 Sanger expressed sequence tags (EST), and 1,260 full-length cDNA sequences. This resulted in 917 scaffolds and an N50 of 706 kbp (Supplementary Table S1). The size of the *P.*

*capsici* genome is intermediate among the sequenced oomycetes, larger than *Albugo* spp. and *Pythium ultimum* (up to 45 Mbp) (Kemen et al. 2011; Levesque et al. 2010) but more compact than *Phytophthora ramorum*, *P. sojae*, *Hyaloperonospora arabidopsidis*, and *P. infestans* at 65, 95, 100, and 240 Mbp, respectively (Baxter et al. 2010; Haas et al. 2009; Tyler et al. 2006). The *P. capsici* draft genome had fewer scaffolds or a larger scaffold N50, which may reflect hybrid assembly or fewer repeat sequences in *P. capsici* (19%) than *P. infestans*, *P. sojae*, or *P. ramorum* (74, 39, and 28%, respectively) (Haas et al. 2009; Tyler et al. 2006). Like other *Phytophthora* spp., the majority of *P. capsici* repeats were retrotransposons (84%) with long terminal repeats, of which 57% were Gypsy elements. The mitochondrial genome (80,148 bp) was assembled in four scaffolds.

Gene model prediction using a training set of 16.4 million sequencing-by-synthesis mRNA-sequences from nine life stages



**Fig. 1.** Life cycle. **A**, Thick-walled sexual oospores are produced when plants are infected with both the A1 and A2 mating types. **B**, Oospores germinate to initiate the asexual life cycle. **C** and **G**, Millions of deciduous sporangia are produced on the surface of infected plants. **D**, Sporangia can germinate and cause infection directly or **E**, germinate in the presence of free water to produce 20 to 40 bi-flagellate swimming zoospore that swim to plants and **F**, cause infection directly.

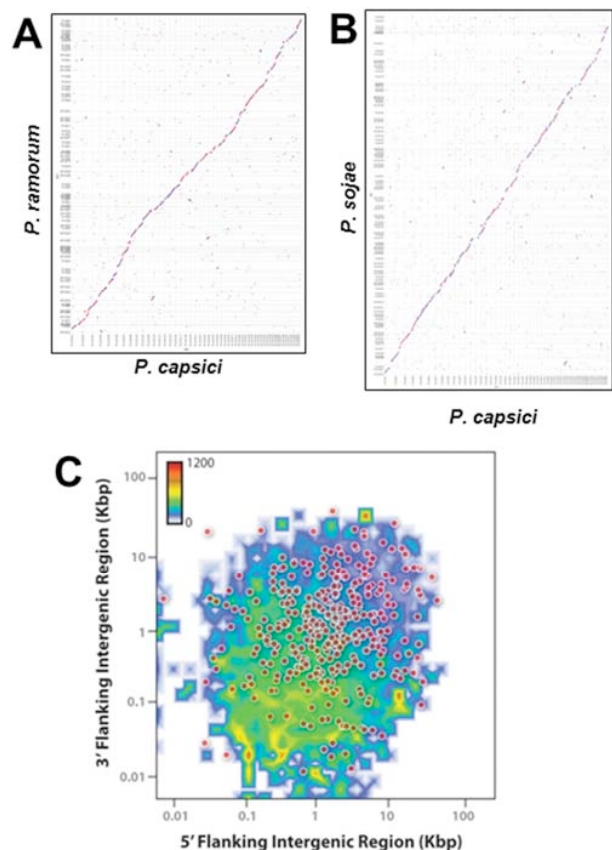
predicted 19,805 genes supported by cDNA sequences (50%) or previously identified proteins (75%) (Supplementary Tables S2 and S3). After excluding 2,682 genes with homology to transposable elements, *P. capsici* had a gene content similar to that of *P. infestans*, *P. sojae*, and *P. ramorum* (17,797, 16,988 and 14,451, respectively) (Haas et al. 2009; Tyler et al. 2006) and had slightly more core eukaryotic genes identified compared with these other species (Supplementary Fig. S2). The nonrepetitive gene density in *P. capsici* (268/Mbp) was higher than other *Phytophthora* spp. (74/Mbp in *P. infestans*, 179/Mbp in *P. sojae*, and 222/Mbp in *P. ramorum*) (Haas et al. 2009; Tyler et al. 2006).

### Comparison with other *Phytophthora* spp.

Genome comparisons revealed almost perfect scaffold-level synteny between the gene models of *P. capsici*, *P. ramorum*, and *P. sojae*, with very little duplication (Fig. 2A and B). There was also extensive synteny with *P. infestans* genes (Supplementary Fig. S3). The genome organization of *P. capsici* was similar to other *Phytophthora* spp., with two-thirds of the genes located in gene-rich blocks that have a conserved order

and few repeats (median intergenic distance of 351 bp), separated by relatively gene-poor regions without order conservation and more repetitive sequences (median intergenic distance approximately 3kb) (Fig. 2C) (Haas et al. 2009). Like *P. infestans*, the 365 predicted *P. capsici* genes without a homolog resided predominantly in gene-poor genomic regions (Fig. 2C) (Haas et al. 2009; Raffaele et al. 2010).

In all, 78% of the predicted *P. capsici* genes clustered into 2,483 candidate multigene families. Among these families, there are cytoplasmic effector genes of the RxLR type, including homologs of known avirulence genes in *P. infestans* (Boutemy et al. 2011; Dou et al. 2008; Whisson et al. 2007). *P. capsici* had 357 RxLR effector genes, which is similar to the numbers reported for *P. sojae* and *P. ramorum*, whereas *P. infestans* had >500 (Haas et al. 2009; Tyler et al. 2006). *P. capsici* also had 29 putative full-length genes and 70 pseudogenes for crinkling and necrosis (CRN) effectors, an ancient class of intracellular effector proteins that share a highly conserved N-terminal domain, required for translocation, and diverse C-terminal effector domains (Haas et al. 2009). Although markedly fewer than in *P. infestans*, the majority of predicted CRN effector domains were conserved between species.



**Fig. 2.** Synteny and genome architecture of *Phytophthora capsici*. Pairwise comparison of gene models from *P. capsici* to **A**, *P. ramorum* and **B**, *P. sojae*. Sequences other than gene models were replaced with Ns. Maximal unique matches in all six frames were used as anchors for amino-acid-based alignment with the PROmer package of MUMmer. Scaffolds >450 kbp are shown (*P. capsici*: 46 Mb, *P. ramorum*: 36 Mb, *P. sojae*: 58 Mb). Scaffolds were ordered to maximize the center diagonal. There were some translocations, which possibly were assembly errors. A small region of *P. sojae* (top) had no syntenic genes in *P. capsici*. **C**, *P. capsici* whole-genome architecture illustrated by the distribution of all predicted genes according to the length of their 5' (x axis) and 3' (y axis) intergenic regions, counted by two-dimensional binning. The color scale shows number of genes in bins. In addition, the 365 predicted proteins for which no homolog was found ( $10^{-5}$  E-value cutoff) are indicated as dots.

### Single-nucleotide variation and genetic mapping.

Nucleotide diversity was assessed in seven *P. capsici* field isolates (including parental isolates TN1 and TN2) and an isolate of the sister species *P. tropicalis* by restriction-site-associated DNA (RAD) sequencing (Supplementary Table S4) (Baird et al. 2008). Between 5 and 9 million sequencing-by-synthesis reads, flanked by a *SgrAI* restriction site, were generated per isolate (Supplementary Table S5). The reads were aligned to the *P. capsici* reference genome and covered approximately 2.3 Mbp at a depth of  $\geq 30$ -fold (Baranzini et al. 2010; Bell et al. 2011; Kim et al. 2009). The single-nucleotide variant (SNV) density among *P. capsici* isolates was 23.4/kb (range of 16.9 to 25.9/kb) and between *P. capsici* and the closely related species *P. tropicalis* was 51.7/kb (Supplementary Tables S6 and S7; Supplementary Figs. S4 and S5). This is much higher than the SNV diversity observed among *P. infestans* strains (0.65/Kb average) and between *P. infestans* T30-4 and species in its clade (7.67/Kb average) (Raffaele et al. 2010). The nucleotide diversity ( $\pi$ ) was 0.012 (range of 0.009 to 0.016), approximately 16-fold higher than *Homo sapiens* and approximately fourfold higher than obtained by RAD in natural stickleback populations and most outbred plant species. The three-spined stickleback (*Gasterosteus aculeatus* L.) is an outcrossing species of fish studied to better understand adaptive evolution (Hohenlohe et al. 2010).

In addition, the inheritance of SNV markers was examined in 65 putative  $F_1$  intercross progeny by *PstI*-based RAD (Supplementary Fig. S6; Supplementary Table S8). Of these 65 isolates, 2 were found to be clonally derived from the TN2 parent and 3 additional isolates were clonal to progeny isolates. The remaining 60 true  $F_1$  progeny were genotyped at bi-allelic sites exhibiting either a di- or tri-modal distribution, where the peaks corresponded to homozygous reference alleles, heterozygotes, and homozygous alternate alleles (Supplementary Fig. S7) (Kim et al. 2009). Genotypes were imputed at 500,352 nucleotides, with  $\geq 10$  uniquely aligned reads in  $\geq 55$  progeny. Homozygosity was inferred when >90% reads displayed one allele, and heterozygosity when 20 to 80% of reads displayed one allele. We have previously shown such genotype imputation to be >99% accurate (Baranzini et al. 2010; Bell et al. 2011; Kim et al. 2009). In total, 45% (9,254) of SNV markers showing Mendelian inheritance were in coding regions, and 3,382 (37%) of these were nonsynonymous (Supplementary Fig. S8a; Supplementary Table S10). The distribution of muta-



tions in *Phytophthora* clade 1c species is in a similar range, with 35% of the SNV found in coding regions, 41% of which are nonsynonymous (Raffaele et al. 2010). This was surprising because there is typically extensive purifying selection on genes. A parsimonious explanation of these findings is that rates of mutation relative to generation times are much greater in *P. capsici* than higher eukaryotes.

In total, 20,568 SNV were either polymorphic in the parents and had Mendelian segregation ( $\chi^2 \geq 0.05$ ) or were homozygous for opposite alleles in the parents and heterozygous in progeny that did not exhibit LOH (marker spacing 3,111 bp) (Supplementary Table S9). A genetic map was developed using markers on the 108 scaffolds that had sufficient SNV densities for independent genetic linkage analysis in the F<sub>1</sub> progeny. Of these, 92 scaffolds had SNV markers inherited as a single co-segregating block, while markers on 15 scaffolds broke into two blocks and markers on the largest scaffold (scaffold 1) broke into three blocks (Supplementary Table S11; Supplementary Fig. S9). Scaffold-level linkage groups (LG) were inherited with log of the likelihood ratio scores of  $\geq 7$ . Following scaffold-level linkage analysis, a genetic map of *P.*

*capsici* was constructed with 1,136 SNV markers having segregation patterns  $<95\%$  similar. The genetic map was 1,654 centimorgans (cM) in length and comprised 18 LG accounting for 84% (54 Mbp) of the reference genome and containing approximately 90% of the predicted genes (Fig. 3; Supplementary Table S12). The order for the majority of markers was the same on the genetic and physical maps. Genetic linkage analysis localized the mating type locus to an approximately 300-kb critical region on scaffold block 4.1 in LG10.

### LOH.

Integrated genomic and genetic analysis allowed assessment of chromosomal variation among the 65 putative F<sub>1</sub> intercross progeny. In all, 23 F<sub>1</sub> isolates and two clonal isolates contained regions with LOH. LOH was identified by homozygosity at all markers in two or more adjacent (linked) RAD sequences that were incompatible with parental genotypes (Supplementary Fig. S10; Supplementary Table S13). The length of LOH tracts varied from 299 bp to  $>1$  Mbp. LOH occurred in 14 of the 18 LG and in 59 scaffolds or scaffold blocks. In total, 54% (11,048) of the SNV markers exhibited LOH in at least one isolate and

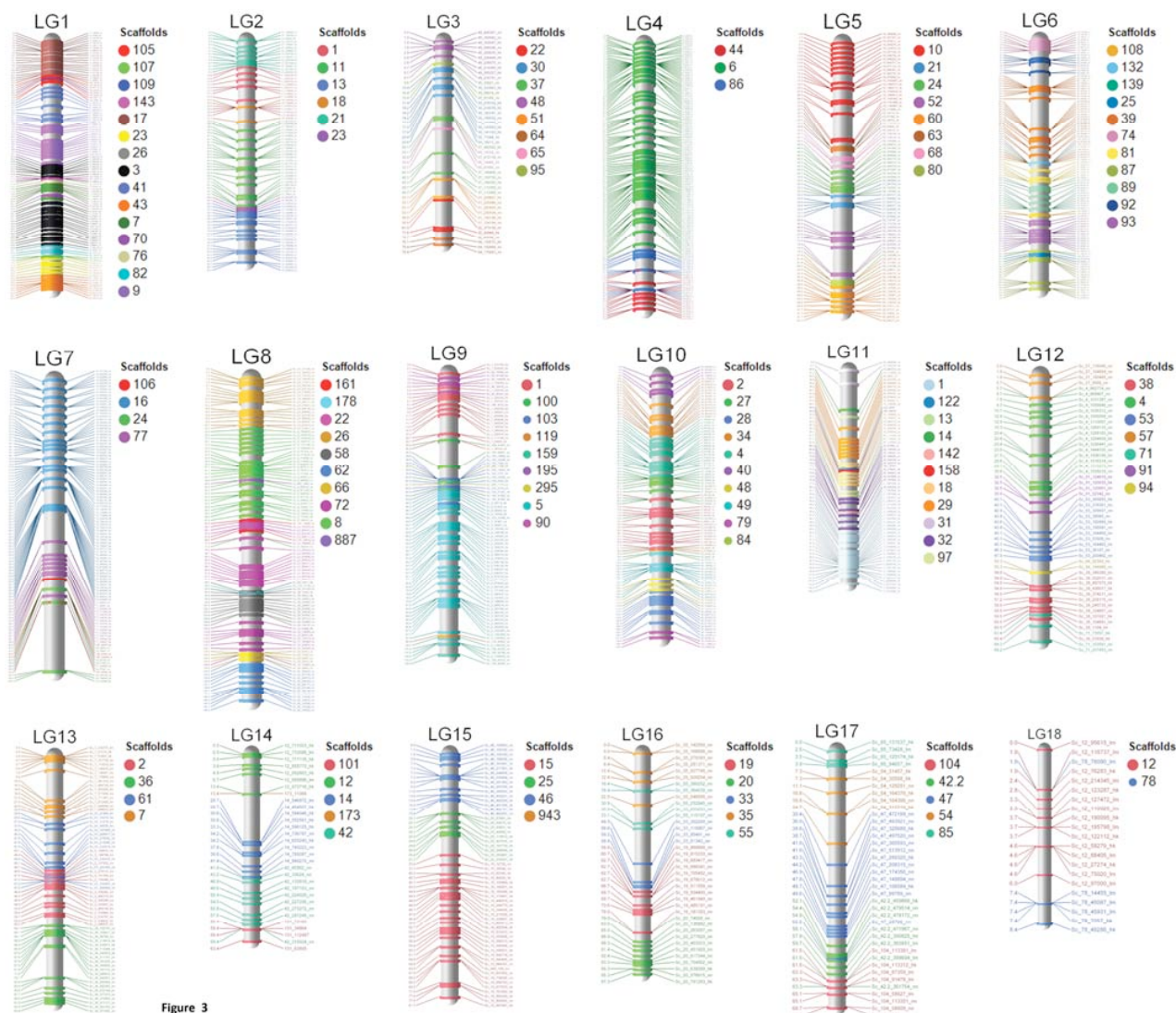


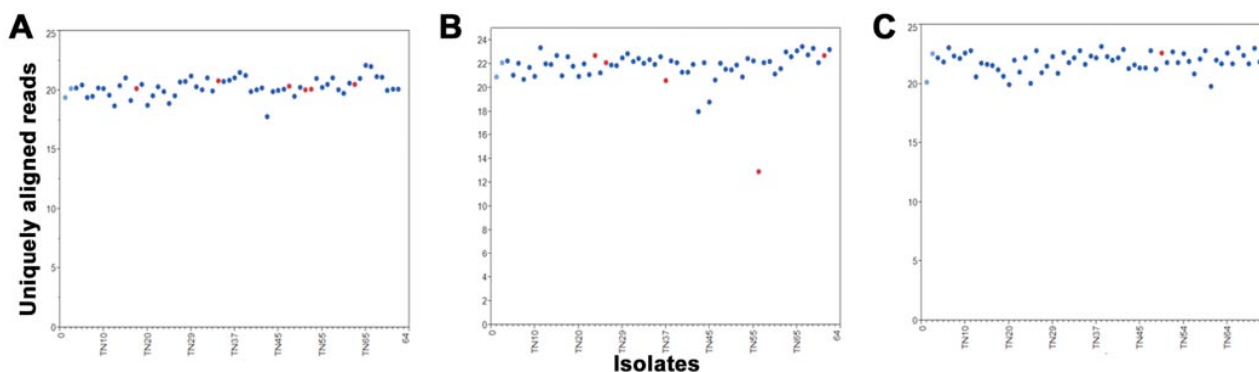
Figure 3

**Fig. 3.** The 1,654-centimorgan *Phytophthora capsici* genetic map. The map contains 1,136 single-nucleotide variant markers and 18 linkage groups (LG). LG are colored by genome scaffold.

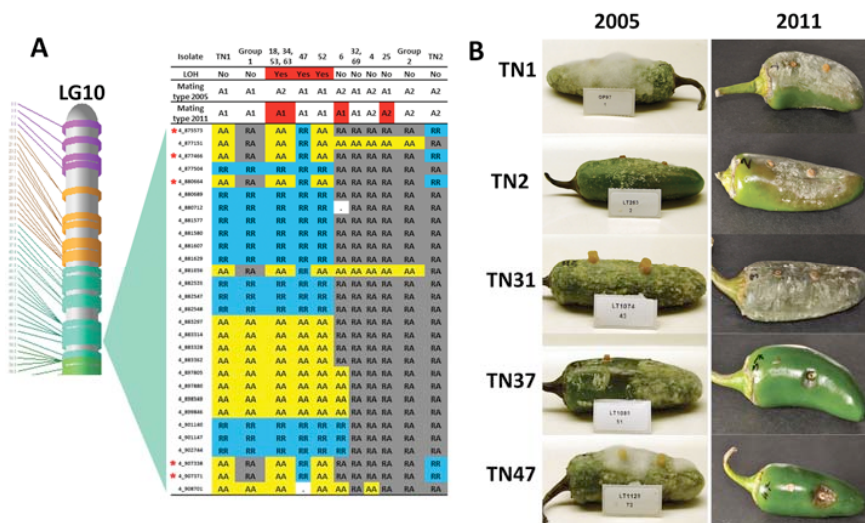
the LOH was bi-directional. The frequency and distribution of LOH tracts varied among isolates and 36.5% (19.7 Mbp) of the genome was affected by LOH in at least one isolate. Two isolates (TN37 and TN47) had LOH exceeding 10% of the genome (Supplementary Table S14). Interestingly, LOH showed bias toward gene-rich regions. Given the requirements of informative parental genotypes, at least two adjacent RAD tags to be affected; and given that LOH was sought only in the mapped genome, the true extent of LOH is likely to be considerably higher.

Subchromosomal mitotic LOH has several molecular mechanisms, including deletions (resulting in hemizyosity) and translocations (resulting in homozygosity), associated with homologous and nonhomologous recombination or gene conversion (Chamnanpant et al. 2001; Cvitanich et al. 2006). Of note, mutagenesis-induced LOH in *P. capsici* was not associated with subchromosomal deletions (Hulvey et al. 2010). In only 1 of

12 isolates was LOH due to hemizyosity as assessed by batched, normalized sequence coverage plots (Bell et al. 2011) (Fig. 4A to C). In all, 45% (4,967) of the SNV markers exhibiting LOH were exonic, 17% (1,911) were in coding regions, and 3.5% (391) were predicted to be nonsynonymous, affecting 264 genes. Phenotypic consequences of genome diversity were sought among F<sub>1</sub> progeny. Five F<sub>1</sub> isolates switched mating type from A2 to A1 between assessments in 2005 and RAD sequencing in 2010, while one switched from A1 to A2 (Fig. 5A). Isolates were in storage for most of this interval and were subcultured <10 times. Four of the isolates that switched from A2 to A1 had >800-Kbp tracts of LOH encompassing a region that appears to contain the mating type locus (or loci) on LG10 (Fig. 5A). TN25, which changed from A1 to A2, did not exhibit LOH at the mating type locus but did have >1.6 Mbp LOH in LG16. However, A1 mating type isolates TN37 and TN70 had similar LOH on LG16 but did not change mat-



**Fig. 4.** Ploidy for loss of heterozygosity (LOH) regions. **A**, Normalized sequence coverage of restriction-site-associated DNA tags across 50,649 nucleotides (nt) of an 850,537-nt LOH tract on scaffold 4 (linkage group [LG]10) at nucleotides 86,143 to 936,679 encompassing the mating type locus. The six isolates with LOH are denoted by red dots. Normalized coverage was obtained by dividing the average coverage across the tract by the total unique coverage for each sample. Parents are in light blue and non-LOH progeny are dark blue. **B**, Normalized coverage of tags across 14,143 nt of a 415,652-nt LOH tract on scaffold 35 (LG16) at nucleotides 141,346 to 556,998. Five isolates with LOH are in red. TN58 had one-half of the expected coverage, indicating LOH to reflect a subchromosomal deletion. **C**, Normalized coverage of tags across 22,124 nt of a 655,528-nt LOH tract on scaffold 1 (LG9) at nucleotides 87,441 to 742,969. TN50 had LOH of this region.



**Fig. 5.** Loss of heterozygosity (LOH) in *Phytophthora capsici*. **A**, Graphical genotypes for linkage group (LG)10 in the region of scaffold 4 encompassing the mating type locus in parents (TN1 and TN2, black) and 59 F<sub>1</sub> progeny. R = reference allele and A = alternate allele. Red asterisks to the left of the marker names denote unambiguous LOH loci. Dots represent missing genotypes. Six progeny had LOH in this region (in red), including four that switched mating type from A2 to A1 and two that remained A1 following LOH. Group 1: progeny 10, 13, 15, 16, 26, 29, 31, 35, 37-41, 46, 51, 54, 59-61, 64-66, 70, 71, and 9. Group 2: progeny 14, 17, 19-24, 27, 30, 33, 36, 42-45, 49, 5, 55, 58, 62, 67, and 68. **B**, Association of LOH with loss of pathogenicity of *P. capsici* isolates in wound-inoculated jalapeno fruit between 2005 and 2011. TN1 and TN2 are the parents. TN31 was representative for all progeny except TN37 and TN47. In 2011, the TN37 and TN47 genomes had 10.4 and 11.1% LOH, respectively, and these isolates no longer infected wounded fruit.

ing type. Thus, LOH at the mating type locus was associated with mating type switch in four of six isolates.

LOH was also associated with a loss of pathogenicity to pepper and cucumber fruit. Between tests in 2005 and 2011, progeny isolates TN37 and TN47 became unable to infect wounded or unwounded fruit (Fig. 5B). In 2011, host cells consistently elicited a strong hypersensitive response (cell necrosis) at point of inoculation, suggesting loss of a factor necessary for infection, such as an Avh protein. This is characteristic of *Phytophthora* Avh proteins, which evolve rapidly in a presumed host–pathogen arms race (Haas et al. 2009; Jiang et al. 2008; Tyler et al. 2006; Whisson et al. 2007). These isolates had similar growth rates in vitro in 2005 and 2011 even after substantial LOH. Indeed, the TN37 and TN47 genomes each had >10% LOH and share an LOH tract on LG1 (17\_31,860 to 17\_870,195 and 105\_51,382 to 105\_63734). A plausible explanation is that LOH changed the expression or expressed sequence of one or more *P. capsici* effectors in the avirulent isolates.

## DISCUSSION

In summary, the genome sequence of *P. capsici* was typical for the genus in organization and gene content. A prototypical, dense genetic map indicated the recombination rate of *P. capsici* to be 38.7 kbp/cM, an order of magnitude less than yeast but 30-fold higher than *H. sapiens* (Li and Freudenberg 2009). SNV density and diversity among *P. capsici* genomes was much higher than in other eukaryotes and coding and nonsynonymous variants were remarkably common. This suggests that germline mutation rates may be much higher in *P. capsici* than in *H. sapiens*, *Saccharomyces cerevisiae*, *Drosophila melanogaster*, and *P. infestans*, or that purifying selection was less (Baranzini et al. 2010). The overall SNV diversity for the seven *P. capsici* field isolates was similar but markedly different from the *P. tropicalis* isolate. The isolates from Peru and Argentina appear to be part of long-lived, widely dispersed clonal lineages where, unlike the situation for the North America isolates, sexual recombination may be rare (Gobena et al. 2012; Hurtado-Gonzales et al. 2008). Interestingly, the isolates from Peru and Argentina share the most polymorphic loci and may be derived from a common, invasive population or widely dispersed clonal lineage. The high level of SNV diversity in *P. capsici* has many practical implications for research, including the ability to finely map segregating phenotypes in larger populations of progeny, conduct very sensitive analyses of population structure and overall dynamics, and track the fate of numerous heterozygous loci during asexual growth (e.g., tracking LOH).

*P. capsici* populations are often characterized by high levels of diversity and rapid adaptation (Erwin and Ribeiro 1996; Gevens et al. 2008; Gobena et al. 2012; Hausbeck and Lamour 2004; Hurtado-Gonzales et al. 2008; Lamour and Hausbeck 2001; Lamour and Kamoun 2009; Lamour et al. 2012; Leonian 1922; Meitz et al. 2010). In addition to elevated genetic diversity, *P. capsici* genomes also had a remarkably high frequency of mitotic LOH occurring more commonly in gene-rich areas. Mitotic genomic instability was also recently described in *P. ramorum* (Vercauteren et al. 2011), suggesting that LOH may be a general characteristic of oomycetes. Here, we show that loss of pathogenicity and mating type switches are likely phenotypic consequences of mitotic LOH. For the latter, the evidence for causality of LOH was strong: four of five progeny that switched from the A2 to the A1 mating type had tracts of LOH spanning the region where mating type maps in this cross. This was reminiscent of the mating type locus in the fungus *Candida albicans*, where a repressor suppresses switch-

ing to the mating-competent form (Morschhauser 2010). Loss of mating type heterozygosity relieves this repression, ensuring that only mating-type-homozygous cells are mating competent. An ability to switch mating type has important consequences, particularly in northern latitudes where winter or fallow-month survival requires thick-walled, sexually produced oospores (Lamour et al. 2012). It also provides a mechanism to introgress mutations efficiently into a population. The fungicide mefenoxam, for example, is strongly inhibitory to *P. capsici* and has been widely used in some areas. Mefenoxam resistance is inherited in *P. capsici* as a single incompletely dominant trait (Lamour and Hausbeck 2000, 2001). Mating type switches in a rare, partially resistant field isolate will double the chances for outcrossing. In addition to aiding introgression of a novel resistance allele by sexual reproduction, LOH could directly catalyze the conversion to full resistance. Using the SNV markers presented here, it should be possible to detect instances of LOH within naturally occurring isolates.

The dramatic expansion of repeated sequences observed in several pathogen lineages among oomycetes (*P. infestans*, *H. arabidopsidis*) and fungi (*Puccinia graminis* f. sp. tritici, *Melampsora larci-populina*, *Blumeria graminis*) was proposed to contribute to fast adaptation (Haas et al. 2009; Spanu et al. 2010). Indeed, a clear correlation was observed between the rate of gene evolution and the amount of flanking repeats (Raffaele et al. 2010). By contrast, the *P. capsici* genome is rather compact, with a relatively low repeat content, and yet shows remarkable plasticity. This plasticity may be due, in part, to LOH, which represents an alternative mechanism for rapidly generating diversity, preferentially affecting gene dense regions. Previous studies suggest a role for LOH in adaptive evolution (Rosenberg 2011). It has, for example, been associated with oncogenesis and embryonic stem cell adaptation in *H. sapiens* (Narva et al. 2010), and fluconazole resistance in *C. albicans* (Forche et al. 2011). Here, we extend these observations, suggesting that the co-occurrence of frequent mitotic LOH and high nucleotide diversity in genes may provide a general mechanism for rapid adaptation in *P. capsici* and may contribute to the difficulties encountered when studying oomycetes outside of natural epidemics (e.g., loss of pathogenicity and virulence) and the success of oomycetes as plant pathogens.

## MATERIALS AND METHODS

### Isolates.

Three in vitro crosses were completed to produce a moderately inbred isolate for sequencing the reference genome (Hurtado-Gonzales and Lamour 2009). Crosses include an F<sub>1</sub> and two recurrent backcrosses (BC1 and BC2). An isolate from the BC2 progeny was chosen for genome sequencing based on growth rate and asexual and sexual spore production (Hurtado-Gonzales and Lamour 2009). The initial cross was between an A1 isolate recovered from cucumber in Michigan in 1997 (TN1) and an A2 isolate recovered from pumpkin in Tennessee in 2004 (TN2). The backcrosses were to the TN2 parent. Crosses were made by co-culturing the parental isolates on V8-juice agar plates, incubating the plates for 8 to 12 weeks in darkness, and then separating the sexual oospores from the subtending asexual mycelium and other asexual spores (sporangia and zoospores) through mechanical shearing and an enzyme treatment to destroy the asexual material (Lamour and Hausbeck 2000). The enzymes not only destroy the asexual material, they also stimulate oospore germination. Germinating oospores were harvested using a 5- $\mu$ l pipette while looking at the spores under a light microscope. Once harvested, the germinated spores are placed onto V8 media, subcultured to produce mycelium for DNA and to conduct

mating type analysis, and then stored long term (3 to 5 years) in water with sterile hemp seed.

The progeny from all of the crosses were genotyped using amplified fragment length polymorphism markers, tested for mating type, assessed for growth rate on agar media, and assayed for pathogenicity and virulence on detached pepper ('Early Jalapeno') and cucumber fruit (various cultivars) (Hurtado-Gonzales and Lamour 2009). All fruit inoculations were conducted on two fruit and the experiments repeated twice. Pathogenicity was assessed by placing a 7-mm plug of mycelium at both a wounded and nonwounded site. Wounded sites are distal and nonwounded sites proximal to the stem end of the fruit. Pathogenicity to pepper fruit was assessed in 2005 and 2011. In total, 69 progeny from the  $F_1$  were further analyzed using a focused resequencing strategy (outlined below). In addition, the mating type was tested at three different time intervals: within 6 months of the cross, at the time of the focused resequencing, and 1 year following the resequencing.

### Genome sequencing and assembly.

High-quality genomic DNA was prepared from the isolate selected for genome sequencing using a phenol/chloroform protocol. Aliquots from the same batch of DNA were submitted to 454 Inc. (Branford, CT, U.S.A.) and to the Joint Genome Institute (JGI, Walnut Creek, CA, U.S.A.) for 454- and Sanger-based sequencing, respectively. In total, 2,451,737 random shotgun genomic reads, 2,022,402 2.3-Kb paired-end reads, and 833,606 18.0-Kb paired-end reads were produced using the 454 Titanium chemistry. In addition, 197,306 random genomic paired-end reads from a 6.3-Kb plasmid library, 146,588 paired-end reads from a 34.7-Kb fosmid library, and 65,000 EST were produced at the JGI using Sanger sequencing. Additionally, 1,260 full-length cDNA sequences were generated by Sanger sequencing and cDNA sequence assembly.

The sequence reads were assembled using a modified version of Arachne v.20071016 (Jaffe et al. 2003) with parameters `maxcliq1 = 100`, `n_haplotypes = 2`, and `BINGE_AND_PURGE = True`. The output of this assembly was passed through Rebuilder and SquashOverlaps with parameters to merge adjacent assembled alternative haplotypes and, subsequently, run through another complete Arachne assembly process to finalize the assembly. Prior to assembly, redundant 454 pairs (defined as at least 98% identical across their length) were removed and a single pair retained. An error correction step was applied to all 454 data that attempted to set base pairs to equal the Sanger sequence. This was accomplished by shredding the Sanger sequence to a size similar to the 454 and then aligning these shredded reads and the 454 simultaneously. Where the 454 read sequence had no support, we changed the 454 read sequence to match the Sanger read sequence. This removed many of the 454-bp errors (primarily insertions and deletions and occasional base-pair substitutions) and coalesced the haplotypes if the Sanger data were all from one haplotype. Error correction was also applied during the Arachne assembly process by screening each scaffold against bacterial proteins, organelle sequences, and GenBank using Megablast against GenBank NR and blastp against a set of known microbial proteins. In all, 28 scaffolds identified as prokaryotic contamination were removed. We classified additional scaffolds as mitochondrion ( $n = 4$ ), small repetitive ( $n = 6$ ), alternative haplotype ( $n = 17$ ), and less than 1 kb ( $n = 1,050$ ). The final assembly consists of 971 scaffold sequences, with half of the genome contained in 29 scaffolds of at least 705.7 kb in length and contigs greater than 34.6 kb. Scaffolds ranged in size from 1 kb to 2.1 Mb with an N50 of 705.7 kb. The genome size was tested using the approach outlined by Baxter with nucleotide-by-nucleotide coverage calculated by aligning back to the refer-

ence (Baxter et al. 2010). This resulted in an estimated genome size of approximately 55 Mbp. We also estimated the genome size by measuring the percentage of all reads that were captured by the assembly using a high-quality 10,000-read subset (average quality  $\geq 35$ , length  $\geq 100$ , and no missing bases) and extrapolating genome size if the assembly had captured all the reads. We extrapolated up from the genome size that doesn't include gaps (56.1 Mb) and 97.44% of the reads were captured, leading to an estimated genome size of approximately 57 Mbp. Both estimates likely underestimate the genome size due to overalignment of repeat copies, possibly explaining the discrepancy between our gapped genome size (64.1 Mb) and these lower estimates of total genome size. Nevertheless, both estimates are similar to the nongapped genome size (56.1 Mb), suggesting that the assembly is missing very little of the nonrepetitive genome.

In addition, total RNA was isolated from different growth conditions (rich media, starvation, and sporulating), pooled, and submitted to both 454 and the JGI for Sanger cDNA and EST sequencing. At the JGI, EST were produced as previously described with the following modification: two cDNA libraries were constructed and sequenced and the size ranges of the cDNA library inserts were 0.6 kb to 2 kb and  $>2$  kb (Jeffries et al. 2007).

### Genome sequence analysis and annotation of gene families.

The genome assembly of *P. capsici* LT1534 was annotated using the JGI Annotation Pipeline, which combines several gene predictors: i) cDNA-based gene models were derived from 1,260 full-length cDNAs and 11,090 consensus sequences clustered from 56,448 EST, then mapped to genomic sequence; ii) protein-based gene models were predicted using FGENESH+ (Salamov and Solovyev 2000) and GeneWise (Birney et al. 2004) seeded by BLASTx alignments of genomic sequence against sequences from the NCBI nonredundant protein set nr; and iii) ab initio gene models were predicted using FGENESH trained on the set of putative full-length genes and reliable protein-based models. GeneWise models were completed using scaffold data to find start and stop codons. EST, EST clusters, EST contigs, and full-length cDNAs were used to verify, complete, and extend the gene models. Because multiple gene models per locus were often generated, a single representative gene model for each locus was chosen based on homology and EST support, and used for further analysis. This led to a filtered set of 19,805 gene models (including 2,682 genes with homology to transposable elements), with their properties and support by different lines of evidence. All gene-based analyses were done on the full 19,805-gene set. In total, 56,448 EST are available in GenBank under the title "DOE Joint Genome Institute *Phytophthora capsici* EST project". An additional 1,260 full-length cDNAs are available in GenBank under the title "DOE Joint Genome Institute *Phytophthora capsici* cDNA project". The genome sequence is available online.

All predicted gene models were functionally annotated by the JGI Annotation Pipeline using InterProScan (Zdobnov and Apweiler 2001) and hardware-accelerated double-affine Smith-Waterman alignments against highly curated databases such as SwissProt (Bairoch et al. 2005), Kyoto Encyclopedia of Genes and Genomes (KEGG) (Ogata et al. 1999), and Pfam (Bateman et al. 2004). KEGG hits were used to map EC numbers (Bairoch 2000), and InterPro, KEGG, and SwissProt hits were used to map gene ontology terms (Ashburner et al. 2000). In addition, predicted proteins were annotated according to KOG classification (Koonin et al. 2004). Protein targeting predictions were made with signalP (Nielsen et al. 1999) and TMHMM (Krogh et al. 2001). Finally, all proteins were aligned by BLASTp to proteins in nr and to each other; after



the latter analysis, the alignment scores were used as a distance metric for clustering by MCL into a first draft of 2,483 candidate multigene families. The same method was used to group genes with those of other *Phytophthora* spp. We used BLASTP to identify CRN coding genes from the *P. capsici* gene model set (Supplementary Table S16). For this purpose, 16 well-characterized CRN proteins from *P. infestans* were searched against the *P. capsici* protein data set. Sequences with significant hits ( $E = <10$  to 5) were retrieved and a nonredundant CRN-like protein set was created (Win et al. 2006). CRN-like protein sequences were manually examined for the presence of LFLAK motifs and C-terminal effector domains. To determine whether candidates are full length or pseudogenes, we aligned our CRN candidates with previously characterized CRN. CRN proteins of at least 250 amino acid residues or more and that aligned with both the N- and C-terminal domains of PiCRN were considered full length (Haas et al. 2009). Finally, we used SignalP3.0 to identify predicted secretion signals for all CRN candidates. Candidate RxLR effectors were mined via the protocol outlined previously (Haas et al. 2009). In brief, all open reading frames (ORF) greater than 210 nucleotides (70 amino acids) from the genome were generated and translated to amino acid coding sequences. Those containing an RxLR (Arg–anything–Leu–Arg) between 30 and 60 amino acids from the start, a signal peptide score greater than 0.9, and a cleavage site prior to the RxLR are considered effectors (Supplementary Table S17).

The CEGMA pipeline was used to compare the completeness and continuity of six Oomycete genomes (Parra et al. 2007). CEGMA checks the completeness and continuity of genome assemblies on the basis of 248 core eukaryotic genes. Here, we compared the CEGMA analyses (without KOG 69) for the genomes of *P. sojae*, *P. ramorum*, *P. infestans*, *Pythium ultimum*, and *Hyaloperonospora arabidopsidis* with the draft genome of *Phytophthora capsici* reported here (Supplementary Table S15).

Transposon-like coding sequences were identified by searching the annotated protein-coding sequences of *P. capsici* against databases of transposon sequences with TransposonPSI. Nucleotide sequences corresponding to the coding regions of the annotated protein coding genes were searched against the TransposonPSI collection of transposon protein sequences using BLASTX. This included ORF derived from mobile elements discovered in other sequenced *Phytophthora* genomes. In addition, the *P. capsici* protein sequences were searched against a collection of PSI-BLAST profiles corresponding to common families of transposable elements, including Gypsy-retrotransposons and Piggybac-transposons. Any *P. capsici*-annotated coding sequence identified by having a BLAST  $E$  value of, at most,  $1e-10$  was identified and reported, yielding 2,682 candidate gene annotations with transposon homology. Transposon family (PSI-BLAST) and top transposon protein BLASTX match are provided.

Scaffolds, gene models clusters, and annotations thereof may be accessed at the JGI *P. capsici* portal (Grigoriev et al. 2012).

### Focused resequencing.

RAD sequencing was employed to identify and genotype single-base substitutions in 7 field isolates (including the parents TN1 and TN2) and 69 progeny (TN3 to TN71) of the F<sub>1</sub> cross (Baird et al. 2008). Briefly, genomic DNA was digested with the 6-bp recognition restriction endonuclease *Pst*I or the 8-bp recognition *Sgr*AI (Florigenex, Inc., Eugene, OR, U.S.A.). The digested DNA was sheared to between 200 and 500 bp and the adaptors were ligated to the ends which contained sequencing priming sites and an isolate-specific 5-bp

tag (Baird et al. 2008). Between 5 and 9 million sequencing-by-synthesis reads of 72 nucleotides in length were generated per isolate using an Illumina GAI sequencer. The short reads abutting the restriction site are referred to as RAD tags. The RAD approach reduces complexity by limiting sequences to regions directly adjacent (left and right) to a restriction site, and the isolate-specific 5-bp tags allow multiplexed sequencing. Reads were aligned to the *P. capsici* reference, covering approximately 2 million nucleotides at a depth of  $\geq 30$ -fold (Baranzini et al. 2010; Bell et al. 2011; Kim et al. 2009).

### SNV discovery and genotyping.

Alignments of the RAD sequence and subsequent variant detection were carried out using the Genomic Short-read Nucleotide Alignment Program (Wu and Nacu 2010) and the Alpheus pipeline (Miller et al. 2008), as described (Baranzini et al. 2010; Bell et al. 2011; Kim et al. 2009). Reads were aligned to the *P. capsici* genome assembly (version 11) and alignments retained that matched the genome with 95% identity or higher and that had five or fewer equally high-scoring alignments. Reads with equally good hits to more than one genomic region, reflecting repeat content or redundancy in the assembly, were not considered uniquely aligned. Four isolates were excluded from analysis due to insufficient aligned reads. Read-count-based allele frequencies of bi-allelic nucleotide variants had a tri-modal distribution, with peaks corresponding to homozygous reference alleles, heterozygotes, and homozygous alternate alleles (Kim et al. 2009).

Potential segregating SNV sites were identified in the parents and progeny using the following criteria:  $\geq 10\times$  unique sequence coverage, an average quality score  $\geq 20$ , and an alternate allele frequency  $\geq 20\%$  within unique reads. This pool of potential segregating sites was then analyzed across all lines and SNV genotyping criteria were developed based on a visual assessment of the distribution of alternate alleles at different levels of unique sequence coverage. SNV were genotyped for all lines with at least  $10\times$  unique coverage where alternate allele frequency within unique reads was  $<10\%$  (homozygous for the reference allele),  $>20\%$  and  $<80\%$  (heterozygous), or  $>90\%$  (homozygous for the alternate allele). Our final pool of markers was limited to loci with five or fewer missing genotypes in the progeny. This approach does not enrich or bias for gene-rich areas of the genome.

Nucleotide density and diversity was calculated for SNV sites in the *Sgr*AI RAD tags from the *P. capsici* parents and five additional unrelated isolates (two of which represent very large clonal populations in Peru and Argentina) and a *P. tropicalis* isolate. Heterozygosity and homozygosity were distinguished by allele frequency cutoffs of 15 and 85% and there was no requirement for number of missing genotypes in the isolates. SNV diversity was calculated as previously described (Begun et al. 2007).

### LOH and clonal isolates.

LOH was assessed using graphical genotypes constructed by ordering the SNV loci according to i) the genetic linkage of markers in the individual scaffolds and ii) the genetic linkage of markers from all scaffolds together (see below). The minimum tracts of LOH were calculated for regions with continuous, uninterrupted switching of loci to either the reference or alternate allele (homozygosity) in a way that is impossible by normal meiosis or clonal reproduction (Judd and Petes 1988). To avoid minor genotyping errors, LOH was only assigned for tracts where the switched loci spanned at least two different sequencing sites. In addition, the junction of LOH versus non-LOH in an isolate with good sequence coverage was analyzed to compare the coverage on either side of the junction.



Genetic similarity of the parents and progeny was assessed in JMP Genomics by analyzing a relationship matrix for our final pool of markers using a Fast Ward Hierarchical clustering at the default settings. Alleles were assumed to be identical by descent. Putative clones were confirmed using the graphical genotypes. Genetic similarity analyses revealed three clonal lineages. One clonal lineage contained the parent isolate TN2 and the (mistaken) F<sub>1</sub> progeny TN3 and TN50. Two other clonal lineages were derived from true F<sub>1</sub> progeny and included isolates TN56, TN57, and TN63 in one clonal lineage and isolates TN8 and TN11 in another clonal lineage (Supplementary Fig. S11). This was not surprising because occasional sporangia survive the enzymatic treatment used to destroy asexual propagules, and premature oospore germination can produce sporangia that appear to be discrete oospore progeny. A single isolate of each lineage was retained for genetic linkage analysis. Apomixis, which occurred in other crosses stemming from these parents, is another possible explanation (Hurtado-Gonzales and Lamour 2009).

### Genetic linkage analysis.

Putative segregating markers were analyzed using a  $\chi^2$  test for simple Mendelian inheritance (5% significance). In total, 20,568 SNV markers were either polymorphic in the parents and had Mendelian segregation ( $\chi^2 \geq 0.05$ ) or were homozygous for opposite alleles in the parents and heterozygous in progeny that did not exhibit LOH. These were used for genetic linkage analysis (average marker spacing 3,111 bp).

In total, 3,141 SNV markers exhibited segregation ratios <5% by  $\chi^2$  test. Some of these clustered in blocks and may reflect gene conversion, as described in other *Phytophthora* spp. (Chamnanpant et al. 2001; Cvitanich et al. 2006). An additional approximately 100 loci were heterozygous in both parents and all progeny and may reflect assembly artifacts, copy number variants, or obligate heterozygosity.

Mendelian markers were further analyzed using JoinMap 4.1 at the default settings for a cross heterogeneously heterozygous with phase unknown (van Ooijen 2011). Initially, markers from individual scaffolds were analyzed to determine whether they are inherited as a linked unit (expected if the scaffold is assembled correctly). If a scaffold broke into more than one piece, the pieces are referred to as scaffold blocks. Following this, markers from all the scaffolds, or scaffold blocks, were analyzed together to produce a linkage map. The LG were further refined to include only markers with segregation patterns <95% similar and to exclude isolates showing any LOH within the LG. Markers are named with the scaffold number followed by the nucleotide position on the scaffold. Additional information was included for the markers used in JoinMap 4.1 to designate the three possible segregation scenarios: *hk* = heterozygous in both parents, *nn* = parent 1 is homozygous and parent 2 is heterozygous, and *lm* = parent 1 is heterozygous and parent 2 is homozygous.

The SNV genotyping error rates were assessed for both parents (TN1 and TN2) at 95 polymorphic sites in portions of 30 genes across 14 LG by Sanger sequencing of polymerase chain reaction products. Of these 190 genotypes, all but five were confirmed. The five genotypes that differed occurred on a single gene that switched to homozygosity at all five sites.

### ACKNOWLEDGMENTS

This work was supported by grants to S. F. Kingsmore from the National Science Foundation (05-512), United States Department of Agriculture, and JGI Community Sequencing Program. The work conducted by the United States Department of Energy JGI is supported by the Office of Science of the United States Department of Energy under contract number DE-AC02-05CH11231. J. Mudge was supported, in part, by NIH grant

RR016480. S. Raffaele was supported by a Marie Curie Fellowship (contract 255104). R. Stam and E. Huitema were supported by the BBSRC and Royal Society of Edinburgh. L. M. Cano, J. Win, and S. Kamoun were supported by the Gatsby Charitable Foundation. R. Sharma and M. Thines are supported by the LOEWE Initiative of the state of Hesse. K. H. Lamour was supported by a grant from the UTIA AgResearch program. We thank P. Richardson and E. Rubin for JGI support in genome sequencing. Author contributions consisted of manuscript preparation by K. H. Lamour, S. F. Kingsmore, J. Mudge, S. Kamoun, E. Huitema, S. Raffaele, I. V. Grigoriev, A. Kuo, R. Stam, A. K. Bharti, and M. Thines; crosses, phenotypic analyses, and genetic analyses by K. H. Lamour, D. Gobena, O. P. Hurtado-Gonzales, A. Savidor, J. Hulvey, and D. Storey; genomic or transcriptomic sequencing by J. R. Knight, J. P. Affourtit, E. A. Lindquist, S. F. Kingsmore, J. Mudge, D. L. Dinwiddie, C. Detter, O. Chertkov, and C. S. Han; assembly by J. Schmutz, D. Platt; annotation by A. Kuo, A. Salamov, and I. V. Grigoriev; alignments and variant detection by N. A. Miller, B. J. Rice, J. Mudge, and K. H. Lamour; and analysis by K. H. Lamour, J. Mudge, S. F. Kingsmore, S. Kamoun, E. Huitema, D. Storey, S. Raffaele, L. M. Cano, A. Kuo, I. V. Grigoriev, A. K. Bharti, R. S. Donahoo, S. Finley, A. Savidor, D. Gobena, O. P. Hurtado-Gonzales, J. Hulvey, R. Sharma, R. Stam, D. Storey, M. Thines, J. Win, B. J. Haas, C. Detter, C. S. Han, and O. Chertkov.

### LITERATURE CITED

- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. 2000. Gene ontology: Tool for the unification of biology. *The Gene Ontology Consortium. Nat. Genet.* 25:25-29.
- Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., Cresko, W. A., and Johnson, E. A. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3:e3376. Published online.
- Bairoch, A. 2000. The ENZYME database in 2000. *Nucleic Acids Res.* 28:304-305.
- Bairoch, A., Apweiler, R., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M. J., Natale, D. A., O'Donovan, C., Redaschi, N., and Yeh, L. S. 2005. The Universal Protein Resource (UniProt). *Nucleic Acids Res.* 33:D154-D159.
- Baranzini, S. E., Mudge, J., van Velkinburgh, J. C., Khankhanian, P., Khrebukova, I., Miller, N. A., Zhang, L., Farmer, A. D., Bell, C. J., Kim, R. W., May, G. D., Woodward, J. E., Caillier, S. J., McElroy, J. P., Gomez, R., Pando, M. J., Clendenen, L. E., Ganusova, E. E., Schilkey, F. D., Ramaraj, T., Khan, O. A., Huntley, J. J., Luo, S., Kwok, P. Y., Wu, T. D., Schroth, G. P., Oksenberg, J. R., Hauser, S. L., and Kingsmore, S. F. 2010. Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis. *Nature* 464:1351-1356.
- Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., and Sonnhammer, E. L. L. 2004. The Pfam protein families database. *Nucleic Acids Res.* 32:D138-D141.
- Baxter, L., Tripathy, S., Ishaque, N., Boot, N., Cabral, A., Kemen, E., Thines, M., Ah-Fong, A., Anderson, R., and Badejoko, W. 2010. Signatures of adaptation to obligate biotrophy in the *Hyaloperonospora arabidopsidis* genome. *Science* 330:1549.
- Begun, D. J., Holloway, A. K., Stevens, K., Hillier, L. W., Poh, Y. P., Hahn, M. W., Nista, P. M., Jones, C. D., Kern, A. D., Dewey, C. N., Pachter, L., Myers, E., and Langley, C. H. 2007. Population genomics: Whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *PLoS Biol.* 5:e310. Published online.
- Bell, C. J., Dinwiddie, D. L., Miller, N. A., Hateley, S. L., Ganusova, E. E., Mudge, J., Langley, R. J., Zhang, L., Lee, C. C., Schilkey, F. D., Sheth, V., Woodward, J. E., Peckham, H. E., Schroth, G. P., Kim, R. W., and Kingsmore, S. F. 2011. Carrier testing for severe childhood recessive diseases by next-generation sequencing. *Sci. Transl. Med.* 3:65ra64.
- Birney, E., Clamp, M., and Durbin, R. 2004. GeneWise and Genomewise. *Genome Res.* 14:988-995.
- Boutemy, L. S., King, S. R., Win, J., Hughes, R. K., Clarke, T. A., Blumenschein, T. M., Kamoun, S., and Banfield, M. J. 2011. Structures of *Phytophthora* RXLR effector proteins: A conserved but adaptable fold underpins functional diversity. *J. Biol. Chem.* 286:35834-35842.
- Chamnanpant, J., Shan, W. X., and Tyler, B. M. 2001. High frequency mitotic gene conversion in genetic hybrids of the oomycete *Phytophthora sojae*. *Proc. Natl. Acad. Sci. U.S.A.* 98:14530-14535.
- Cvitanich, C., Salcido, M., and Judelson, H. S. 2006. Concerted evolution of a tandemly arrayed family of mating-specific genes in *Phytophthora* analyzed through inter- and intraspecific comparisons. *Mol. Genet. Genomics* 275:169-184.

- Dou, D., Kale, S. D., Wang, X., Chen, Y., Wang, Q., Jiang, R. H., Arredondo, F. D., Anderson, R. G., Thakur, P. B., McDowell, J. M., Wang, Y., and Tyler, B. M. 2008. Conserved C-terminal motifs required for avirulence and suppression of cell death by *Phytophthora sojae* effector Avr1b. *Plant Cell* 20:1118-1133.
- Erwin, D. C., and Ribeiro, O. K. 1996. *Phytophthora* Diseases Worldwide. The American Phytopathological Society, St. Paul, MN, U.S.A.
- Forche, A., Abbey, D., Pisithkul, T., Weinzierl, M. A., Ringstrom, T., Bruck, D., Petersen, K., and Berman, J. 2011. Stress alters rates and types of loss of heterozygosity in *Candida albicans*. *MBio* 2:e00129-11. Published online.
- Gevens, A. J., Donahoo, R. S., Lamour, K. H., and Hausbeck, M. K. 2008. Characterization of *Phytophthora capsici* causing foliar and pod blight of snap bean in Michigan. *Plant Dis.* 92:201-209.
- Gobena, D., Roig, J., Galmarini, C., Hulvey, J., and Lamour, K. H. 2012. Genetic diversity of *Phytophthora capsici* isolates from pepper and pumpkin in Argentina. *Mycologia* 104:102-107.
- Grigoriev, I. V., Nordberg, H., Shabalov, I., Aerts, A., Cantor, M., Goodstein, D., Kuo, A., Minovitsky, S., Nikitin, R., Ohm, R. A., Otiillar, R., Poliakov, A., Ratnere, I., Riley, R., Smirnova, T., Rokhsar, D., and Dubchak, I. 2012. The genome portal of the department of energy joint genome institute. *Nucleic Acids Res.* 40:D26-D32.
- Haas, B. J., Kamoun, S., Zody, M. C., Jiang, R. H., Handsaker, R. E., Cano, L. M., Grabherr, M., Kodira, C. D., Raffaele, S., Torto-Alalibo, T., Bozkurt, T. O., Ah-Fong, A. M., Alvarado, L., Anderson, V. L., Armstrong, M. R., Avrova, A., Baxter, L., Beynon, J., Boevink, P. C., Bollmann, S. R., Bos, J. I., Bulone, V., Cai, G., Cakir, C., Carrington, J. C., Chawner, M., Conti, L., Costanzo, S., Ewan, R., Fahlgren, N., Fischbach, M. A., Fugalstaff, J., Gilroy, E. M., Gnerre, S., Green, P. J., Grenville-Briggs, L. J., Grifith, J., Grunwald, N. J., Horn, K., Horner, N. R., Hu, C. H., Huitema, E., Jeong, D. H., Jones, A. M., Jones, J. D., Jones, R. W., Karlsson, E. K., Kunjeti, S. G., Lamour, K., Liu, Z., Ma, L., Maclean, D., Chibucos, M. C., McDonald, H., McWalters, J., Meijer, H. J., Morgan, W., Morris, P. F., Munro, C. A., O'Neill, K., Ospina-Giraldo, M., Pinzon, A., Pritchard, L., Ramsahoye, B., Ren, Q., Restrepo, S., Roy, S., Sadanandom, A., Savidor, A., Schornack, S., Schwartz, D. C., Schumauer, U. D., Schwessinger, B., Seyer, L., Sharpe, T., Silvar, C., Song, J., Studholme, D. J., Sykes, S., Thines, M., van de Vondervoort, P. J., Phuntumart, V., Wawra, S., Weide, R., Win, J., Young, C., Zhou, S., Fry, W., Meyers, B. C., van West, P., Ristaino, J., Govers, F., Birch, P. R., Whisson, S. C., Judelson, H. S., and Nussbaum, C. 2009. Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature* 461:393-398.
- Hausbeck, M. K., and Lamour, K. H. 2004. *Phytophthora capsici* on vegetable crops: Research progress and management challenges. *Plant Dis.* 88:1292-1303.
- Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., and Cresko, W. A. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet.* 6:e1000862. Published online.
- Hulvey, J., Young, J., Finley, L., and Lamour, K. 2010. Loss of heterozygosity in *Phytophthora capsici* after N-ethyl-nitrosourea mutagenesis. *Mycologia* 102:27-32.
- Hurtado-Gonzales, O. P., and Lamour, K. H. 2009. Evidence for inbreeding and apomixis in close crosses of *Phytophthora capsici*. *Plant Pathol.* 58:715-722.
- Hurtado-Gonzales, O., Aragon-Caballero, L., Apaza-Tapia, W., Donahoo, R., and Lamour, K. 2008. Survival and spread of *Phytophthora capsici* in coastal Peru. *Phytopathology* 98:688-694.
- Jaffe, D. B., Butler, J., Gnerre, S., Mauceli, E., Lindblad-Toh, K., Mesirov, J. P., Zody, M. C., and Lander, E. S. 2003. Whole-genome sequence assembly for mammalian genomes: Arachne 2. *Genome Res.* 13:91-96.
- Jeffries, T. W., Grigoriev, I. V., Grimwood, J., Laplaza, J. M., Aerts, A., Salamov, A., Schmutz, J., Lindquist, E., Dehal, P., Shapiro, H., Jin, Y. S., Passoth, V., and Richardson, P. M. 2007. Genome sequence of the lignocellulose-bioconverting and xylose-fermenting yeast *Pichia stipitis*. *Nat. Biotechnol.* 25:319-326.
- Jiang, R. H., Tripathy, S., Govers, F., and Tyler, B. M. 2008. RXLR effector reservoir in two *Phytophthora* species is dominated by a single rapidly evolving superfamily with more than 700 members. *Proc. Natl. Acad. Sci. U.S.A.* 105:4874-4879.
- Judd, S. R., and Petes, T. D. 1988. Physical lengths of meiotic and mitotic gene conversion tracts in *Saccharomyces cerevisiae*. *Genetics* 118:401-410.
- Kemen, E., Gardiner, A., Schultz-Larsen, T., Kemen, A. C., Balmuth, A. L., Robert-Seilantantz, A., Bailey, K., Holub, E., Studholme, D. J., Maclean, D., and Jones, J. D. 2011. Gene gain and loss during evolution of obligate parasitism in the white rust pathogen of *Arabidopsis thaliana*. *PLoS Biol.* 9:e1001094. Published online.
- Kim, J. I., Ju, Y. S., Park, H., Kim, S., Lee, S., Yi, J. H., Mudge, J., Miller, N. A., Hong, D., Bell, C. J., Kim, H. S., Chung, I. S., Lee, W. C., Lee, J. S., Seo, S. H., Yun, J. Y., Woo, H. N., Lee, H., Suh, D., Lee, S., Kim, H. J., Yavartanoo, M., Kwak, M., Zheng, Y., Lee, M. K., Park, H., Kim, J. Y., Gokumen, O., Mills, R. E., Zaranek, A. W., Thakuria, J., Wu, X., Kim, R. W., Huntley, J. J., Luo, S., Schroth, G. P., Wu, T. D., Kim, H., Yang, K. S., Park, W. Y., Kim, H., Church, G. M., Lee, C., Kingsmore, S. F., and Seo, J. S. 2009. A highly annotated whole-genome sequence of a Korean individual. *Nature* 460:1011-1015.
- Koonin, E. V., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Krylov, D. M., Makarova, K. S., Mazumder, R., Mekhedov, S. L., Nikolskaya, A. N., Rao, B. S., Rogozin, I. B., Smirnov, S., Sorokin, A. V., Sverdlov, A. V., Vasudevan, S., Wolf, Y. I., Yin, J. J., and Natale, D. A. 2004. A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol.* 5:R7.
- Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. 2001. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* 305:567-580.
- Lamour, K., and Kamoun, S. 2009. Oomycete genetics and genomics: Diversity, interactions, and research tools. Blackwell Publishing, Hoboken, NJ, U.S.A.
- Lamour, K. H., and Hausbeck, M. K. 2000. Mefenoxam insensitivity and the sexual stage of *Phytophthora capsici* in Michigan cucurbit fields. *Phytopathology* 90:396-400.
- Lamour, K. H., and Hausbeck, M. K. 2001. The dynamics of mefenoxam insensitivity in a recombining population of *Phytophthora capsici* characterized with amplified fragment length polymorphism markers. *Phytopathology* 91:553-557.
- Lamour, K. H., Stam, R., Jupe, J., and Huitema, E. 2012. The oomycete broad-host-range pathogen *Phytophthora capsici*. *Mol. Plant Pathol.* 13:329-337.
- Leonian, L. H. 1922. Stem and fruit blight of peppers caused by *Phytophthora capsici* sp. nov. *Phytopathology* 12:400-408.
- Levesque, C. A., Brouwer, H., Cano, L., Hamilton, J. P., Holt, C., Huitema, E., Raffaele, S., Robideau, G. P., Thines, M., and Win, J. 2010. Genome sequence of the necrotrophic plant pathogen *Pythium ultimum* reveals original pathogenicity mechanisms and effector repertoire. *Genome Biol.* 11:R73.
- Li, W., and Freudenberg, J. 2009. Two-parameter characterization of chromosome-scale recombination rate. *Genome Res.* 19:2300-2307.
- Meitz, J. C., Linde, C. C., Thompson, A., Langenhoven, S., and McLeod, A. 2010. *Phytophthora capsici* on vegetable hosts in South Africa: Distribution, host range and genetic diversity. *Australas. Plant. Pathol.* 39:431-439.
- Miller, N. A., Kingsmore, S. F., Farmer, A., Langley, R. J., Mudge, J., Crow, J. A., Gonzalez, A. J., Schilkey, F. D., Kim, R. J., van Velkinburgh, J., May, G. D., Black, C. F., Myers, M. K., Utsey, J. P., Frost, N. S., Sugarbaker, D. J., Bueno, R., Gullans, S. R., Baxter, S. M., Day, S. W., and Retzel, E. F. 2008. Management of high-throughput DNA sequencing projects: Alpheus. *J. Comput. Sci. Syst. Biol.* 1:132.
- Morschhauser, J. 2010. Regulation of white-opaque switching in *Candida albicans*. *Med. Microbiol. Immunol.* 199:165-172.
- Narva, E., Autio, R., Rahkonen, N., Kong, L., Harrison, N., Kitsberg, D., Borghese, L., Itskovitz-Eldor, J., Rasool, O., Dvorak, P., Hovatta, O., Otonkoski, T., Tuuri, T., Cui, W., Brustle, O., Baker, D., Maltby, E., Moore, H. D., Benvenisty, N., Andrews, P. W., Yli-Harja, O., and Lahesmaa, R. 2010. High-resolution DNA analysis of human embryonic stem cell lines reveals culture-induced copy number changes and loss of heterozygosity. *Nat. Biotechnol.* 28:371-377.
- Nielsen, H., Brunak, S., and von Heijne, G. 1999. Machine learning approaches for the prediction of signal peptides and other protein sorting signals. *Protein Eng.* 12:3-9.
- Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., and Kanehisa, M. 1999. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 27:29-34.
- Parra, G., Bradnam, K., and Korf, I. 2007. CEGMA: A pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23:1061-1067.
- Raffaele, S., Farrer, R. A., Cano, L. M., Studholme, D. J., MacLean, D., Thines, M., Jiang, R. H., Zody, M. C., Kunjeti, S. G., Donofrio, N. M., Meyers, B. C., Nussbaum, C., and Kamoun, S. 2010. Genome evolution following host jumps in the Irish potato famine pathogen lineage. *Science* 330:1540-1543.
- Rosenberg, S. M. 2011. Stress-induced loss of heterozygosity in *Candida*: A possible missing link in the ability to evolve. *MBio* 2:e00200-11. Published online.
- Salamov, A. A., and Solov'yev, V. V. 2000. *Ab initio* gene finding in *Drosophila* genomic DNA. *Genome Res.* 10:516-522.
- Spanu, P. D., Abbott, J. C., Amselem, J., Burgis, T. A., Soanes, D. M., Stuber, K., Ver Loren van Themaat, E., Brown, J. K., Butcher, S. A., Gurr, S. J., Lebrun, M. H., Ridout, C. J., Schulze-Lefert, P., Talbot, N. J.,

- Ahmadinejad, N., Ametz, C., Barton, G. R., Benjdia, M., Bidzinski, P., Bindschedler, L. V., Both, M., Brewer, M. T., Cadle-Davidson, L., Cadle-Davidson, M. M., Collemare, J., Cramer, R., Frenkel, O., Godfrey, D., Harriman, J., Hoede, C., King, B. C., Klages, S., Kleemann, J., Knoll, D., Koti, P. S., Kreplak, J., Lopez-Ruiz, F. J., Lu, X., Maekawa, T., Mahanil, S., Micali, C., Milgroom, M. G., Montana, G., Noir, S., O'Connell, R. J., Oberhaensli, S., Parlange, F., Pedersen, C., Quesneville, H., Reinhardt, R., Rott, M., Sacristan, S., Schmidt, S. M., Schon, M., Skamnioti, P., Sommer, H., Stephens, A., Takahara, H., Thordal-Christensen, H., Vigouroux, M., Wessling, R., Wicker, T., and Panstruga, R. 2010. Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science* 330:1543-1546.
- Tyler, B. M., Tripathy, S., Zhang, X., Dehal, P., Jiang, R. H., Aerts, A., Arredondo, F. D., Baxter, L., Bensasson, D., Beynon, J. L., Chapman, J., Damasceno, C. M., Dorrance, A. E., Dou, D., Dickerman, A. W., Dubchak, I. L., Garbelotto, M., Gijzen, M., Gordon, S. G., Govers, F., Grunwald, N. J., Huang, W., Ivors, K. L., Jones, R. W., Kamoun, S., Krampis, K., Lamour, K. H., Lee, M. K., McDonald, W. H., Medina, M., Meijer, H. J., Nordberg, E. K., Maclean, D. J., Ospina-Giraldo, M. D., Morris, P. F., Phuntumart, V., Putnam, N. H., Rash, S., Rose, J. K., Sakihama, Y., Salamov, A. A., Savidor, A., Scheuring, C. F., Smith, B. M., Sobral, B. W., Terry, A., Torto-Alalibo, T. A., Win, J., Xu, Z., Zhang, H., Grigoriev, I. V., Rokhsar, D. S., and Boore, J. L. 2006. *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. *Science* 313:1261-1266.
- van Ooijen, J. 2011. Multipoint maximum likelihood mapping in a full-sib family of an outbreeding species. *Genet. Res. (Camb.)* 93:343-349.
- Vercauteren, A., Boutet, X., D'Hondt, L., Bockstaele, E. V., Maes, M., Leus, L., Chandelier, A., and Heungens, K. 2011. Aberrant genome size and instability of *Phytophthora ramorum* oospore progenies. *Fungal Genet. Biol.* 48:537-543.
- Whisson, S. C., Boevink, P. C., Moleleki, L., Avrova, A. O., Morales, J. G., Gilroy, E. M., Armstrong, M. R., Grouffaud, S., Van West, P., and Chapman, S. 2007. A translocation signal for delivery of oomycete effector proteins into host plant cells. *Nature* 450:115-118.
- Win, J., Kanneganti, T. D., Torto-Alalibo, T., and Kamoun, S. 2006. Computational and comparative analyses of 150 full-length cDNA sequences from the oomycete plant pathogen *Phytophthora infestans*. *Fungal Genet. Biol.* 43:20-33.
- Wu, T. D., and Nacu, S. 2010. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* 26:873-881.
- Zdobnov, E. M., and Apweiler, R. 2001. InterProScan—An integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17:847-848.

## AUTHOR-RECOMMENDED INTERNET RESOURCES

- DeCypher TimeLogic biocomputing system: [www.timelogic.com](http://www.timelogic.com)  
 JGI genome portal *Phytophthora capsici* webpage: [jgi.doe.gov/capsici](http://jgi.doe.gov/capsici)  
 Micans MCL website: [www.micans.org/mcl](http://www.micans.org/mcl)  
 TransposonPSI analysis tool: [transposonpsi.sf.net](http://transposonpsi.sf.net)