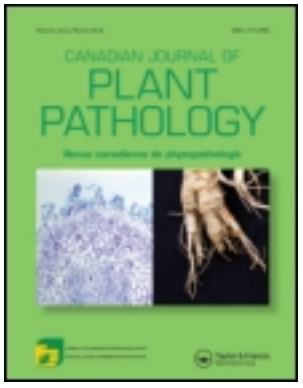

Canadian Journal of Plant Pathology
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tcjp20

Development of a SNP genetic marker system based on variation in microsatellite flanking regions of Phytophthora infestans

C. L. Abbott [a], S. R. Gilmore [a], C. T. Lewis [a], J. T. Chapados [a], R. D. Peters [b], H. W. Platt [b], M. D. Coffey [c] & C. A. Lévesque [a]

[a] Agriculture and Agri-Food Canada, Biodiversity, K.W. Neatby Building, Central Experimental Farm, 960 Carling Avenue, Ottawa, ON K1A 0C6, Canada

[b] Agriculture and Agri-Food Canada, Crops and Livestock Research Centre, 440 University Avenue, Charlottetown, PEI C1A 4N6, Canada

[c] Department of Plant Pathology and Microbiology, University of California, Riverside, CA, 92521, USA

Version of record first published: 22 Sep 2010.

To cite this article: C. L. Abbott , S. R. Gilmore , C. T. Lewis , J. T. Chapados , R. D. Peters , H. W. Platt , M. D. Coffey & C. A. Lévesque (2010): Development of a SNP genetic marker system based on variation in microsatellite flanking regions of Phytophthora infestans , Canadian Journal of Plant Pathology, 32:4, 440-457

To link to this article: http://dx.doi.org/10.1080/07060661.2010.510648

Can. J. Plant Pathol. (2010), **32(4)**: 440–457

Taylor & Francis
Taylor & Francis Group

Genomics/Génomique

Development of a SNP genetic marker system based on variation in microsatellite flanking regions of *Phytophthora infestans*

C. L. ABBOTT[1]*, S. R. GILMORE[1]*, C. T. LEWIS[1], J. T. CHAPADOS[1], R. D. PETERS[2], H. W. PLATT[2], M. D. COFFEY[3] AND C. A. LÉVESQUE[1]

[1]*Agriculture and Agri-Food Canada, Biodiversity, K.W. Neatby Building, Central Experimental Farm, 960 Carling Avenue, Ottawa, ON K1A 0C6, Canada*
[2]*Agriculture and Agri-Food Canada, Crops and Livestock Research Centre, 440 University Avenue, Charlottetown, PEI C1A 4N6, Canada*
[3]*Department of Plant Pathology and Microbiology, University of California, Riverside, CA 92521, USA*

(Accepted 10 June 2010)

Abstract: A single nucleotide polymorphism (SNP)-based molecular marker system for the oomycete pathogen causing late blight on solanaceous hosts, *Phytophthora infestans*, was developed by identifying sequence polymorphisms in microsatellite flanking regions (MFRs). MFRs were identified using the complete genome sequence for this pathogen and SNP rates were assessed by sequencing a total of 14 000 bases from 32 MFRs across a diverse international panel of 32 isolates. SNP rates were highly variable among loci, with nucleotide diversity ranging from a low of zero to a high of 0.013 (one change per 78 bp) and an average across all scorable loci (n = 28) of 0.0023 (one change per 426 bp). Nucleotide diversity estimates are highly conservative due to strict objective scoring rules used when calling SNPs. Overall, 102 SNPs were scored using objective criteria (versus 167 as scored by eye), and could distinguish all isolates in the panel, with pairwise per cent similarity between isolates varying from 58.5% to 99.4%. Observed and expected heterozygosities at most SNP sites were similar. The utility of a 'straincode' for strain typing based on up to 19 unlinked SNPs was investigated. The probability of identity was very low ($PI = 3.5 \times 10^{-6}$) when all 19 SNPs were included; however, to achieve probabilities of identity of about 1% and 0.1%, only four and seven loci were needed, respectively. The approach needs to be assessed with larger panels of isolates, but results suggest that MFRs and SNPs are a largely untapped source of useful sequence polymorphisms for strain typing of microbes and that their utility could extend to ecological and evolutionary research, especially with high throughput sequence-based technologies.

Keywords: fungal diversity, late blight, molecular diagnostics, SNP discovery, strain typing

Résumé: Une méthode faisant appel à un marqueur moléculaire basé sur le polymorphisme d'un nucléotide simple (SNP), conçue pour l'identification de l'agent pathogène oomycète causant le mildiou sur les solanacées hôtes, *Phytophtora infestans*, a été développée en déterminant les séquences polymorphes des régions adjacentes aux microsatellites (MFR). Les MFR ont été identifiées en utilisant la séquence génomique entière de l'agent pathogène. Par la suite, les taux de SNP ont été évalués en séquençant 14 000 bases provenant de 32 MFR issus eux-mêmes d'un panel mondial et diversifié de 32 isolats. Les taux de SNP variaient considérablement d'un locus à l'autre, affichant une diversité nucléotidique oscillant de 0 à 0.013 (un changement par 78 paires de bases (pb)) pour une moyenne de 0.0023 (un changement par 426 pb) chez tous les loci analysés (n = 28). Les estimations de la diversité nucléotidique sont très prudentes étant donné la rigueur et l'objectivité des règles appliquées lors de la lecture des SNP. En tout, 102 SNP ont été évalués en fonction de critères objectifs (versus 167 évalués « à l'œil ») et permettaient de distinguer tous les isolats du panel, avec une similitude du pourcentage des paires chez les isolats variant de 58.5 % à 99.4 %. Les hétérozygoties observées et anticipées à la plupart des sites SNP étaient semblables. L'utilité d'un « code souche » pour le typage des souches basé sur au plus 19 SNP non liés a été étudiée. Lorsque les 19 SNP étaient inclus, la probabilité d'identité était très faible ($PI = 3.5 \times 10^{-6}$). Toutefois, pour atteindre des probabilités d'identité d'environ 1 % et 0.1 %, seulement quatre et sept loci étaient requis,

Correspondence to: C. André Lévesque. E-mail: andre.levesque@agr.gc.ca
*Current address: Department of Fisheries and Oceans, Pacific Biological Station, 3190 Hammond Bay Road, Nanaimo, BC V9T 6N7, Canada.

ISSN 0706-0661 print/ISSN 1715-2992 online
DOI: 10.1080/07060661.2010.510648

respectivement. L'approche doit être évaluée avec de plus grands panels d'isolats, mais les résultats suggèrent que les MFR et les SNP sont des sources largement inexploitées de séquences polymorphes permettant le typage des souches de microbes. De plus, ils pourraient servir en recherche sur l'écologie et l'évolution, particulièrement à l'aide de techniques de séquençage à haut débit.

Mots clés: découverte des SNP, diagnostics moléculaires, diversité fongique, mildiou, typage de souche

Introduction

Virtually all aspects of plant pathology research rely on the ability to accurately identify the etiological agent responsible for the disease under investigation (McCartney *et al.*, 2003). Over recent years, traditional, visually based diagnostic methods for non-human pathogens (e.g. using selective media and microscopy) have been superseded by new molecular diagnostic approaches (as reviewed by Schaad & Frederick, 2002; Lievens & Thomma, 2005; Lévesque, 2007). As the ability to genetically identify microbial species becomes routine, the emphasis moves towards distinguishing among intraspecies units and, ultimately, among individual strains.

For the past two decades, microsatellite DNA markers (also called 'simple sequence repeats' or SSRs) have been the genetic marker of choice for deciphering genetic relationships below the species level. Several features of microsatellite repeat arrays make them ideal for this, including: (i) high within-population and within-individual variability; (ii) co-dominance; and (iii) a wide distribution throughout the genome (reviewed by Li *et al.*, 2002). They have been used extensively in molecular ecology studies on plant and animal populations for well over a decade, and are more recently becoming popular for investigating fungal populations (Sullivan & Faeth, 2004; Banke & McDonald, 2005; Stukenbrock *et al.*, 2006; Zhou *et al.*, 2007). In terms of their potential discriminatory power, microsatellite markers have been used to distinguish among fungal species, including oomycetes (Moon *et al.*, 1999; Lee & Moorman, 2008), subspecies (Smith *et al.*, 2001), morphotypes (Burgess *et al.*, 2001), and strains (Enkerli *et al.*, 2001; Hennequin *et al.*, 2001; Lees *et al.*, 2006).

For the oomycete genus *Phytophthora*, microsatellite data have been used to characterize intraspecific genetic diversity in a few species, namely: *P. cinnamomi* Rands (Dobrowolski *et al.*, 2003); *P. ramorum* Werres, De Cock & Man in 't Veld (Ivors *et al.*, 2006); and the species under study here, *P. infestans* (Mont.) de Bary (Vargas *et al.*, 2009). *Phytophthora infestans* is an oomycete (and thus diploid) pathogen that causes late blight on a range of solanaceous hosts and is best known for its devastating effect on potato and tomato crops. Populations of this pathogen outside parts of South and Central America were initially found to be clonal or otherwise low in genetic diversity, but new genotypes were later detected that arose through a combination of migration and sexual reproduction after the A2 mating type spread into Europe and beyond, starting in 1984 (Fry *et al.*, 1993; Goodwin & Drenth, 1997; Cooke & Lees, 2004). It has been a model species for developing new approaches to genotypic and phenotypic typing of plant pathogens (Fry *et al.*, 1992; Goodwin *et al.*, 1992; Forbes *et al.*, 1998; Cooke & Lees, 2004); genotyping methods used include mtDNA haplotypes based on restriction enzyme digests (Griffith & Shaw, 1998) and length variation in microsatellite alleles (Knapova & Gisi, 2002; Lees *et al.*, 2006; Guo *et al.*, 2009).

Extensive microsatellite genotyping of *P. infestans* associated with the EUCABLIGHT project has seen almost 20 000 isolates genotyped from European populations (Cooke *et al.*, 2009). Aims of this massive undertaking include investigating population change over time and identifying evolutionary forces driving observed changes, as relating these to disease management strategies is expected to improve their chances of success (Cooke *et al.*, 2009). Resolving power of the 11 polymorphic loci used is high; however, setbacks associated with standardizing methods and scoring across laboratories hindered early attempts to compile multiple data sets (Cooke *et al.*, 2009). Indeed, microsatellite alleles are known for being difficult to score due to stutter bands caused by slippage of the polymerase during PCR amplification and the difficulties in digitizing and comparing fragment length information. Thus highly specific technical expertise is required before microsatellite genotyping can be implemented, and there is a strict requirement of consistency when making subjective scoring decisions, which makes standardization across labs difficult at best. In contrast, single nucleotide polymorphisms (SNPs, described below) can be genotyped simply and accurately using real-time PCR (see Mhlanga & Malmberg, 2001) or sequencing, which are standard methods in contemporary molecular genetics labs and generate results that are easily digitized and searchable.

Another detraction of microsatellites can be size homoplasy caused by convergent evolution, parallel evolution, or back mutations in the repeat array itself; or insertion/deletion mutations in the sequence bordering the repeat (Primmer & Ellegren, 1998). In theory, size homoplasy is expected to be especially problematic when

investigating pathogenic microbes prone to human-mediated translocation because their establishment in non-native geographical areas may result from single introduction events. To properly track evolutionary progress following such a severe bottleneck, the ideal scenario is one in which new mutations in the bottlenecked population are highly likely to generate new alleles. This is not the case for microsatellite repeat arrays; new alleles typically differ from ancestral alleles by plus or minus one repeat unit (Kimmel *et al.*, 1996; Ellegren, 2004) and are therefore likely to regenerate an allele length, and potentially a genotype, that is already present in the originating population. Indeed, the high mutational instability of microsatellite repeat units coupled with their propensity for undergoing homoplasious changes is thought to present a real risk of misdiagnoses for fungal species (cf. McEwen *et al.*, 2000).

Single nucleotide polymorphisms (SNPs) are widely used as genetic markers in linkage and association studies and are thought to hold substantial, but as yet largely untapped, potential for addressing ecological and evolutionary questions (but see Bensch *et al.*, 2002; Belfiore *et al.*, 2003; Seddon *et al.*, 2005; Delmotte *et al.*, 2008; Narum *et al.*, 2008). SNPs are co-dominant, single-locus, biallelic markers that result from point mutations leading to single base-pair differences between chromosome sequences (Brumfield *et al.*, 2003). Advantages of SNPs include that they are ubiquitous and abundant in the genome, evolve according to simple mutational models, are not prone to homoplasy, and are easy to genotype (Mhlanga & Malmberg, 2001) and score (Morin *et al.*, 2004). However, because of low polymorphism at SNP loci there are concerns that isolating a sufficient number of loci is simply unrealistic given that conventional analytical methods in evolutionary genetics assume independence among marker loci (Glaubitz *et al.*, 2003). Several methods have been used for finding SNPs, including: sequencing orthologous genomic regions in related species (Bensch *et al.*, 2002; Primmer *et al.*, 2002; Aitken *et al.*, 2004; Seddon *et al.*, 2005); *in silico* comparisons of sequence data from multiple individuals and/or species (Smith *et al.*, 2005); and random sequencing (Fisher *et al.*, 1999; Primmer *et al.*, 2002; Giresse *et al.*, 2007). Studies that isolated SNPs using a single whole genome sequence, as opposed to aligning several genomic sequences from across several individuals or species, are virtually non-existent (but see Adams *et al.*, 2006).

The genome sequences of a few *Phytophthora* species are now available (Tyler *et al.*, 2006; Haas *et al.*, 2009) and they have been used to develop microsatellite markers in *Phytophthora* species (Garnica *et al.*, 2006), including *P. infestans* (Schena *et al.*, 2008). In the current study, we evaluate whether the advantages of SNP-based molecular markers can be married with the advantages of high mutation rates at microsatellite loci by developing a SNP-based marker system for *P. infestans* using sequence polymorphisms in microsatellite flanking regions (MFRs). MFRs are located immediately adjacent to microsatellite repeat arrays and are known to contain both point and length mutations (see Ishibashi *et al.*, 1996; Zardoya *et al.*, 1996; Grimaldi & Crouau-Roy, 1997; Jones *et al.*, 1998; Brohede & Ellegren, 1999); however, they have rarely been targeted as sources of molecular markers (but see Mogg *et al.*, 2002; Primmer *et al.*, 2002; Ablett *et al.*, 2006; Fernández *et al.*, 2008). Targeting SNPs in MFRs could help overcome the hurdle of isolating a sufficient number of SNP loci for ecological studies by ensuring that SNP-rich regions are specifically targeted during the SNP discovery stage. The aims of this study were to: (i) describe an automated SNP discovery method based on *in silico* searches of a complete genome sequence for MFRs; (ii) evaluate levels of sequence polymorphisms in MFRs of *P. infestans* by screening a diverse panel of isolates; and (iii) determine whether SNPs in MFRs are suitable markers for strain typing and evolutionary research.

Materials and methods

Microsatellite identification and primer design

The complete genome sequence for *P. infestans* strain T30-4 was downloaded from the Broad Institute website (http://www.broad.harvard.edu/annotation/genome/phytophthora_infestans/Home.html). The analysis was performed on a pre-release of the currently available data (phytophthora_infestans_0.fasta.zip) which have been recently published (Haas *et al.*, 2009). Microsatellites were identified using the MISA – MIcroSAtellite identification tool (available from http://pgrc.ipk-gatersleben.de/misa/). The minimum targeted microsatellite size was 18 bp for di- and tri-nucleotide repeats and 20 bp for tetra- and penta-nucleotide repeats (misa.ini properties: definition: 2-9 3-6 4-5 5-4, interruptions: 100).

A small pilot experiment was done to determine whether SNPs were indeed present in MFRs of *P. infestans*. An average of 12 isolates collected from Eastern Canada (Prince Edward Island and New Brunswick) in 2006 were sequenced at each of four MFRs to assess SNP rates. A much larger data set of MFR sequences was generated for two of these isolates (DAOM 238591 and DAOM 238592) during development of the final SNP based molecular marker set (described below). Using techniques described previously (Peters *et al.*, 1998,

1999), these isolates were determined to be of the A2 mating type, resistant to metalaxyl-m and displaying an allozyme banding pattern at the glucose phosphate isomerase locus consistent with that of the US-8 genotype, a strain that has come to dominate populations of the pathogen in Canada after its introduction in the mid 1990s (Peters *et al.*, 2001).

Primers were designed to amplify a 1.0–1.2 kb product containing the microsatellite using Primer3 (Rozen & Skaletsky, 2000). MISA includes a script to create a Primer3 input file, which we modified to allow user specifiable Primer3 parameters. We settled on PRIMER_PRODUCT_SIZE_RANGE=1000–200, PRIMER_GC_CLAM P=1, PRIMER_OPT_TM=65, PRIMER_MIN_TM=63, PIMER _ MAX_TM=67. The generated Primer3 input file was post-processed to ensure that the amplified regions contained 500–600 bp both upstream and downstream from the microsatellite repeat itself. This was accomplished by increasing the target size in the Primer3 input file; the target start was moved back 0.5 kb and 1.0 kb was added to the length.

To remove duplicate loci and loci containing apparent internal repeats, the amplicon sequences from the Primer3 output were used as query sequences for a BLAST search of the whole genome sequence. Duplicate loci were identified as those with more than one hit with e-value 0.0; repetitive loci were identified as those with overlapping high scoring sequence pairs within the hit, and both were flagged for removal. Removing duplicate loci is critically important because it ensures that heterozygous sites result from allelic differences at a single locus as opposed to differences among multiple, repeated genomic regions that are co-amplified during PCR. 'Microsatellite families' occur in other taxa (Meglecz *et al.*, 2004; Tero *et al.*, 2006) and likely occur in *P. infestans*. Lees *et al.* (2006) found *P. infestans* microsatellites that generated more than two alleles within single strains. A computer program is available to test for repetitive flanking regions among microsatellites (Meglecz, 2007), but it is not designed for identifying repetitive regions in a complete genome sequence as was done here using BLAST.

Development of marker set: PCR optimization and sequencing

PCR amplification of a test set of 95 *P. infestans* microsatellite loci and one positive control was tried using DNA template from the two US-8 *P. infestans* isolates from the Eastern Canada collection mentioned earlier (DAOM 238591, DAOM 238592). PCR reactions were 10 μL in volume and contained 5–10 ng of genomic DNA, 100 μM dNTP, 1 pmol of each primer, and 0.5× (i.e. 0.1 μL) TITANIUM™ Taq DNA Polymerase (Clontech Laboratories, Inc.) in 1× reaction buffer. PCR cycling parameters were: 3 min at 95 °C; a 'touchdown' of 30 s at 95 °C, 30 s at 65 °C to 53 °C (dropping 3 °C per three cycles), 1.5 min at 72 °C; 25 cycles of 30 s at 95 °C, 30 s at 50 °C, 1.5 min at 72 °C; and one final cycle of 3 min at 72 °C. PCR products were visualized on pre-cast, bufferless, 2% agarose gels (E-Gel® *96*, Invitrogen).

Of the 95 microsatellite primer pairs tried, 59 generated single amplification products of the expected size. A second round of PCR, identical to that described already, was done using these 59 primer pairs and the products could be sequenced without a purification step because low concentrations of dNTPs and primers were used in the PCR mix (cf. Allain-Boule *et al.*, 2004). We sequenced 1 or 2 μL of PCR product in both directions using Applied Biosystems BigDye v3.1 sequencing system and 1.6 pmol of primer in a 10 μL reaction volume. Products were run on an ABI 3130*xl* Genetic Analyzer.

In the first sequencing run, about half (n = 29) of the 59 microsatellite-containing fragments generated high quality sequence data in at least one direction. For most loci, sequence reads did not persist through the repeat itself, so internal sequencing primers were designed as close to the repeat as possible to allow for bidirectional sequencing of one or two sides (F or R in locus names) of the targeted microsatellite flanking region (Table 1). To increase the probability of screening independent genomic regions, we first targeted microsatellite repeats on different supercontigs (with one exception, see Results). Second, to reduce the extent of physical linkage among SNPs in the final marker set, we targeted only one of the two flanking regions for most loci, which we selected based on which direction yielded the cleanest sequencing profile. A total of 1019 sequences were deposited in GenBank (HM166412-HM167430). Each sequence is the consensus of the bidirectional sequencing for one strain at one MFR locus.

SNP discovery

Providing new data on the potential information content in MFRs was a major focus of this study, hence we endeavoured to minimize ascertainment bias by screening a very diverse panel of *P. infestans* isolates to estimate marker variability, and assessing SNP rates at each genomic region for which clean sequence data could be obtained (as opposed to selecting loci for screening based on preliminary data on SNP rates). SNPs were identified by screening 32 *P. infestans* isolates collected from across five continents (Table 2). Several measures

Table 1. Details on the microsatellite loci for which one or both flanking regions was analysed for SNPs. The first four digits of the locus name refer to the number of the supercontig it was found on in the whole genome sequence of strain T30-4 that is publicly available. The repeat location refers to position of the repeat within that supercontig. PCR primers are located on either side of the repeat array such that the whole microsatellite and both flanking regions are amplified; internal sequencing primers were used to facilitate bidirectional sequencing of the target flanking region. The scored fragment of each MFR is positioned between the two boldface primers with two exceptions – at loci 3319_1R and 3408_3R a short stretch (< 100 bp) of single stranded data were scored upstream from the forward primer, as indicated in Table 3 by the location of the first two SNPs at these loci.

Microsat. locus	Microsatellite repeat*	Location of repeat in supercontig	MFR name	MFR analysed	Scorable length	PCR primers 5′–3′	Internal sequencing primer name: Sequence 5′–3′
2873_2	$(TG)_{15}$	199623 – 199652	2873_2R	3′	523	2873_2F_for: ACCAAGGTCATTCAAAACCAACC	**2873_2R_fseq: TGCGCTCGTACAAGATGGT**
						2873_2R_rev: AATGGATGAGAACCCAAAGCG	2873_2R_rseq: TGTAAATCGAGCCAGCAC
2908	$(TCT)_9$	136098 – 136124	2908R	3′	488	2908F_for: CGTGTCCAAAGCCGCAAAC	**2908R_fseq: CAGAGACCGAAGAAGAGCCA**
						2908R_rev: CTGCCACCTCTGCGAACTTG	2908R_rseq: TCGGGCAAGCAGAACAA
2947	$(TCGG)_5$	11280 – 11299	2947R	3′	488	2947F_for: CTGCCGCTACGAACAGCAAG	**2947R_fseq: CAACGCGTTACATAGGGGAA**
						2947R_rev: TACTGGTTCCACGCGCAGTC	2947R_rseq: TTGCGGTCGCACTGGTA
2995	$(TCT)_{10}$	68804 – 68833	2995F	3′	415	**2995F_for: TCAAGCTGCGCTTCCATCAC**	2995F_fseq: TTGCGGACGATGTTGAAC
						2995R_rev: TCAACCACACCCAGCGATTC	**2995F_rseq: TGGAAAGAACCGTTACGAAAGA**
3010	$(AGT)_{14}(N)_{10}(GTA)_6$	119518 – 119587	3010R	3′	385	3010F_for: GCCGCCTCGAAGATAAGGAC	**3010R_fseq: CCGCTGTTGCTTTTCTTGTT**
						3010R_rev: GCAGCCCCGAGTGTTATTGTC	**3010R_rseq: GGATTTGTCGAGAATCGTT**
3027	$(CCT)_8$	46329 – 46352	3027R	3′	501	3027F_for: GGACACGGGAAATCTGGGTG	**3027R_fseq: CCCACCGATCACTCCACCT**
						3027R_rev: CGTGGGCAACTACATGACCG	3027R_rseq: TTTGACATTACGTTCGTG
3115	$(CT)_{10}$	43550 – 43569	3115F	5′	483	**3115F_for: GACTATCACTCCGACGCCCG**	3115F_fseq: CCATCGCCGTTCTCAG
						3115R_rev: GTTGCCGAAATGATACGCCC	3115F_rseq: ATGAGGGTGGCTGAAATGA
							3115F_rseq3: CAAAGCCCTAGTTCACACTGA
3157	$(AT)_{12}$	19297 – 19320	3157F	5′	497	**3157F_for: CGCGATGATGTGGAAGATGG**	**3157F_rseq3: TGTAAACCATGCTCTACGATT**
						3157R_rev: CGGTGACAGTGGTGAGTGCC	
3197	$(CGGA)_6$	212012 – 212035	3197F	5′	394	**3197F_for: GTGTGTCGGGTGGAAAAGCC**	3197F_fseq: TAAGTGTACATCTCTAC
						3197R_rev: GTTTTTGTGGCGACCATCGG	**3197F_rseq: GTACTTAAATGCACCGCTCA**

			3197R	3′	444	3197F_for ***	**3197R_fseq: GTTTCCACGCAAAGCACA**
						3197R_rev	3197R_rseq: TCTAATCGACCGTCGCAAT
3200	$(ACTC)_7$	100589 – 100616	3200F	5′	454	3200F_for: TTGCGAAACGCTGGCATAAC	**3200F_fseq: CGCTGGCATAACTTCATTAGAG**
						3200R_rev: TGAGGTGTTGCTCTGGTGCC	3200F_rseq: TGCGCTCGCTAAGTCAACAT
							3200F_rseq2: GAATTGGGGCTCTATCTGAAC
			3200R	3′	302	3200F_for	**3200R_fseq: TTGGCTGCTACTCCATCGT**
						3200R_rev	**3200R_rseq: GAAGACGAAACCAGAGTTG**
							3200R_rseq3: GAGTTGCAAGAAACTGCGA
3210	$(TGCT)_5$	419261 – 419280	3210R	3′	did not align	3210F_for: ACGACGGAGATTTTGCGAGG	3210R_fseq2: CTGGCAATTTGCAAACTGAC
						3210R_rev: TTAACTTGCCGTGTGGCACTG	3210R_rseq: ACACGGGGACCTTCAGCA
3230_1	$(TACA)_6$	150507 – 150530	3230_1F	5′	430	3230_1F_for: AGCAGCAGTCGTCCGCAAC	**3230_1F_fseq: AGGACTGGAGCTATGAGAC**
						3230_1R_rev: CCATCACGCCAAGCTCTGTG	**3230_1F_rseq2: TGATAAGCTGTACAAACATCGT**
			3230_1R	3′	446	3230_1F_for	3230_1R_fseq2: TGCGTACATCGTTCACAATG
						3230_1R_rev	**3230_1R_fseq3: GGACTCGTTGATCTCGCTA**
3252	$(GAGT)_6(N)_{22}(GAGT)_9$	89896 – 89977	3252F	5′	433	**3252F_for: TTCCCCTTCAGACTGCGTCG**	3252F_fseq: CAAGGCCAGACGGTGCT
						3252R_rev: CACTGATGCTGCCTTCGTCC	**3252F_rseq: CCGACGCTAAGTATAACCTGA**
3279	$(ACGG)_5$	115144 – 115163	3279F	5′	426	3279F_for: GGATGGGATGGTGGGGAAAC	**3279F_fseq: TGTGTATATTTATCGAAAGCA**
						3279R_rev: GGCCAGCGACACTGGAAATG	**3279F_rseq: GCAAAATCGCATTATGTGGT**
3318	$(AAG)_9$	68179 – 68205	3318F	5′	334	**3318F_for: ATCACACTCGGACAGCACGC**	3318F_fseq: AGAGTGGAGACTTTGACTG
						3318R_rev: CACTTGCGCCCTGGAATCTC	**3318F_rseq: TCGGCGCAGAATGACAGA**
			3318R	3′	448	3318F_for	**3318R_fseq: AGCCGATCTCGAAACTGGT**
						3318R_rev	3318R_rseq: ATGATCGACGACAAACTTG

(*Continued*)

Table 1. (Continued.)

Microsat. locus	Microsatellite repeat*	Location of repeat in supercontig	MFR name	MFR analysed	Scorable length	PCR primers 5′–3′	Internal sequencing primer name: Sequence 5′–3′
3319_1	$(ACTC)_5$	105829 – 105848	3319_1R	3′	541*	3319_1F_for: CCTACCGCCACCAAGTGGAC	**3319_1R_fseq: CAGCGCGTCAGGAAGAGGT**
						3319_1R_rev: ACCGCTCGGGGTTACTTGC	3319_1R_rseq: CTGATTGATAGCTGCACTTGA
3332	$(AAG)_{14}$	95278 – 95319	3332F	5′	401	**3332F_for: ACTGAGACCACCACGCCACC**	3332F_fseq: TACTCCAAGACGCAACGA
						3332R_rev: TGGCCAGACGAGAGAACGTG	**3332F_rseq2: ATCCTTCTCCTCCTTGTCCA**
3345	$(TGAG)_{10}(GA)_9$	103233 – 103287	3345F	5′	352	**3345F_for: TGGAGATCGGCGTACCAAGG**	3345F_fseq: GTCTAAAGTAGTAGCAACGCT
						3345R_rev: GCCGACCTACAAGACGCACC	**3345F_rseq: GGATGGGATGCTGTGCTTAG**
			3345R	3′	304	3345F_for	**3345R_fseq: CCGATTGCAGCAGTTAGACAT**
						3345R_rev	**3345R_rseq: GATCTTTCGGTAGTACACGAT**
3350_4	$(TC)_{29}$	1948356 – 1948413	3350_4F	5′	372	3350_4F_for: TGTACCCTTGCATCCACACATC	**3350_4F_fseq: GGATCGCAGACTCGTATG**
						3350_4R_rev: TCGGTCGAGTCAACTGTGGG	**3350_4F_rseq: GACCGTACCTCGTAAATGGA**
3350_9	$(AG)_{24}$	2451638 – 2451685	3350_9F	5′	480	**3350_9F_for: TGCGTTATTGTGCCACCCAC**	3350_9F_fseq: GCACAAAATTCGAGGGTCA
						3350_9R_rev: TGGGTCTGTCAGGAAATCGG	**3350_9F_rseq2: GGTCTATTTGCGCTTCTTCT**
			3350_9R	3′	280	3350_9F_for	**3350_9R_fseq: GGTGGGAATGCCAAGCA**
						3350_9R_rev	**3350_9R_rseq: TGCTACAGGCGTTACGAT**
3361_7	$(TC)_{14}(N)_{24}(TC)_{12}$	1102354 – 1102429	3361_7R	3′	415	3361_7F_for: TTTAGTGGCAACGGCAGCAG	**3361_7R_fseq: GGCCTTTTTAAACGTTTGGA**
						3361_7R_rev: CCCGGAATGACGTCGAAAAAC	**3361_7R_rseq: GCATTTCATTCAAACACAG**
3374	$(CT)_{12}$	100577 – 100600	3374R	3′	did not align	3374F_for: CTCGGTCATGTTGGTGGTGG	3374R_fseq: CGCTGCTCTCATGGTTCGT
						3374R_rev: CGAGTGACGCGCTAGGTGAG	3374R_fseq2: GAAACCGAATGCTCTAACTCTCTCC
							3374R_rseq: CTCCGCACGGGAACCTA
							3374R_rseq2: AAGCATCCMTTCACAACTGCGAGC
							3374R_rseq3: TATAGCGTGCAATCTGCGTAGGCGAC

3384	$(CT)_{10}$	949276 – 949295	3384R	3′	512	3384F_for: CCTTCACGGTGCAAATTAAACG	**3384R_fseq: GCAGCGTGAAAGTGTGAAGA**
						3384R_rev: TGATCTCCTCCGCCAGAACC	**3384R_rseq: TGCTCAGCATGATGCCAT**
3390	$(GA)_{20}$	48482 – 48521	3390F	5′	343	3390F_for: CGTAAATGTCACGGTTTCGGTC	**3390F_fseq: CGATAATCTCGATAACTCTAAATG**
						3390R_rev: GCGACAGTGGGTTTGAGTGC	**3390F_rseq: CCACACCAGTACTGCATTTCT**
3408_3	$(CT)_{16}$	244060 – 244091	3408_3R	3′	464**	3408_3F_for: CAAGTGCAAGGCCAAGAACAG	**3408_3R_fseq: CATCAGCGCTAGCGACCT**
						3408_3R_rev: ACGGTGCGTTCGCTTACTCC	**3408_3R_rseq: CGATGTGCGTAATAAGCAA**
3430_8	$(AC)_{10}$	987264 – 987283	3430_8R	3′	423	3430_8F_for: CCATAGGATGCGACTTGGTGC	**3430_8R_fseq: ACCCGGTCGCATACTAATCT**
						3430_8R_rev: GTTTGGCAAGCGAGTCCGAG	**3430_8R_rseq: CCACAGTGCAAAACAAGTC**
3451_4	$(TTAC)_7$	1149723 – 1149750	3451_4F	5′	423	3451_4F_for: CCAAACACTGTGAGCCGAGC	**3451_4F_fseq: CTTAAAGAAAGTGACCAACGA**
						3451_4R_rev: CAGGCGCCAAGACAATCTCC	**3451_4F_rseq: TGAAAACGTGGTGCAATGGT**
3477_2	$(TAAC)_6$	22083 – 22106	3477_2F	5′	423	3477_2F_for: AAAGACACGCGCCTACACTCG	**3477_2F_fseq: TGTCGTATTTTCAACATTGTC**
						3477_2R_rev: CTTGTTCGTGGCGCAATCAG	**3477_2F_rseq: GCGAAACAGCTCTACGAAGT**

Notes: *Scored length at locus 3319_1R begins 133 bp upstream from the 3′ end of 3319_1R_fseq primer.
**Scored length at locus 3408_3R begins 85 bp upstream from the 3′ end of 3319_1R_fseq primer.
***Sequence provided above.

Table 2. Sampling and genotypic information for the panel of 32 *P. infestans* isolates used here for SNP discovery. mtDNA haplotype and RAPD genotype were known prior to this study. The straincode, represented by each line of SNP data for each isolate, was developed here based on 19 unlinked SNPs in MFRs that were scored using objective criteria except for sites shown in grey and italics: these were scored by eye. SNPs in the straincode are ordered by decreasing discriminatory power as determined by per-locus probability of identity estimates (data not shown). The number after the last underscore in the locus name indicates which SNP was used in the straincode at that locus, and corresponds to the 'SNP no.' in Table 3. Bold-type boxes demarcate the two straincodes that were not unique to a single isolate.

											Locus name and SNP scored at each site in straincode																		
Isolate	Alternate ID	DNA Bank	Mating type	[1]mtDNA haplotype	[2]RAPD genotype	[3]RG57 genotype	[4]Isozyme patterns *Gpi*	[4]Isozyme patterns *Pep*	Host*	Collection location	Year sampled	3200F_13	3332F_1	3230_1R_2	3279F_2	2947R_2	3345F_1	3361_7R_1	3408_3R_2	3318F_1	3350_9F_2	3390F_6	3157F_2	3027R_3	3252F_2	3319R_1	3451_4F_1	3197R_1	2908R_2	2995F_1
P10122	US970018 (Fry)	B122	A1	II b	g142	US-17	100/122	100/100	*Le*	New York, USA (NY-1)	1997	*S*	A	R	W	C	A	T	C	*S*	A	T	A	G	T	T	G	G	A	A
P10108	US980066 (Fry)	B123	A1	II b	US-11	US-11	00/100/11	100/100	*Le*	New York, USA (NY-2)	1996	S	A	R	W	C	A	T	-	*S*	A	T	A	G	T	T	G	G	A	A
P10106	US940330 (Fry)	B140	A2	I a	US-7	US-7	100/111	100/100	*St*	Texas, USA (TX)	1994	S	A	G	T	C	A	T	C	S	A	T	A	G	T	T	G	G	A	A
P10127	96-7-3 (Ristaino)	B89	A2	I a	US-18	US-18	100/100	92/100	*Le*	North Carolina, USA (NC)	1996	*K*	R	R	W	C	R	T	Y	S	A	T	A	G	T	-	G	-	A	A
P10123	80787-94L (Kim)	B376	A2	I a	US-15	US-15	100/100	92/100	*Le*	Pennsylvania, USA (PA)	1994	*S*	A	G	W	C	R	T	Y	G	A	T	A	G	T	T	G	G	A	A
P1844	84/E13a (Shaw)	B428	A2	II a	g123				*St*	Egypt (EGY)	1984	C	A	G	W	Y	G	K	Y	G	A	T	A	G	T	T	G	G	A	A
P10113	US970001 (Fry)	B78	A1	I a	g160	US-17	100/122	100/100	*Le*	Florida, USA (FL)	1996	*C*	A	G	T	C	G	T	-	G	A	T	A	G	T	T	G	G	M	A
P7629	Ca65 (Shaw)	A1459	A1	II b	US-6	US-6	100/100	92/100	*Le*	California, USA (CA-1)	1981	Y	R	R	A	Y	R	T	T	G	A	K	A	G	T	T	G	G	A	A
P7035	189 (Goodwin)	B326	A1	II b	US-6	US-6	100/100	92/100	*Le*	California, USA (CA-2)	1989	Y	R	R	A	Y	R	T	C	G	A	K	A	G	T	T	G	G	A	A
T30-4	Sequenced Genome								*St*	Netherlands (NLD-1)	–	-	R	A	W	Y	A	T	Y	S	A	K	R	G	T	T	G	G	A	R
P1415	MR1 (Cohen)	B257	A2	II a	g123				*St*	Israel (ISR)	1984	Y	G	G	W	Y	G	K	T	G	A	T	A	G	T	T	G	G	A	A
P6576	89038 (Davidse)	A1949	A2	II a	g113				*St*	Netherlands (NLD-2)	1989	C	G	-	T	Y	G	T	-	G	A	K	A	R	T	-	G	G	A	A
P10118	CA920008 (Fry)	B139	A2	I a	g161				*St*	British Columbia, Canada (BC)	1996	Y	R	R	T	Y	G	T	T	C	A	T	A	G	T	T	G	G	A	A
P6747	1169 (Goodwin/Fry)	B347	A1	I b	g105		86/100	92/100	*St*	Poland (POL)	~1989	S	G	A	W	C	G	T	T	G	A	K	R	G	T	T	G	G	A	A
P10157	Fay 2A (Gallegly)	B384	A1	I a	g152				*Le*	West Virginia, USA (WV)	1994	S	G	A	W	C	R	T	T	G	A	T	G	G	T	T	G	G	A	A
P10250	H-12/3/02 (Bakonyi)	B218	A2	II a	g173			96/96	*St*	Hungary (HUN-1)	2002	C	R	A	W	Y	G	T	T	G	A	T	-	G	Y	T	G	G	A	A
P6752	606 (Goodwin/Fry)	B59	A2	I b	g128		100/111/12	100/100	*St*	Saltillo, Mexico (MEX-1)	1989	K	G	R	W	C	G	T	T	G	T	T	A	G	T	T	G	G	A	A
P10257	H-24/02 (Bakonyi)	B191	A1	II a	g148			100/100	*Le*	Hungary (HUN-2)	2002	C	R	G	A	C	A	G	T	G	W	T	-	G	T	-	G	S	A	A
P7036	575 (Goodwin/Fry)	B449	A2	I a	g102				*St*	Toluca, Mexico (MEX-2)	~1986	S	A	G	A	C	A	K	T	S	W	T	-	G	T	T	G	G	A	A
P10110	US940504 (Fry)	A1131	A2	II a	g157	US-13	100/100	100/100	*St*	New York, USA (NY-3)	1994	C	G	G	A	Y	G	G	-	C	W	T	R	R	T	-	G	-	A	A
P12017	49-00 (Fillipov)	A1402	A1	II a	–				*St*	Ryazan Region, Russia (RUS-1)	2000	C	G	A	A	Y	R	G	C	G	A	-	R	G	Y	T	G	G	A	A
P1847	83/6 (Shaw)	B342	A1	I a	g102				*St*	Wales, UK (GBR)	1983	C	-	R	A	T	R	G	*Y*	-	W	K	R	-	Y	T	G	G	A	A
P10344	AP12902 (Ahmad)	B252	A1	I a	g168				*St*	Pakistan (PAK)	~2003	C	R	G	A	Y	G	G	Y	G	A	K	A	G	Y	T	G	G	A	A
P9989	12-1997 (Camyon)	B104	A1	II a	g61				*St*	Sweden (SWE-1)	1997	C	A	A	A	T	R	K	T	S	A	K	A	G	Y	T	G	S	A	A
P6629	511 (Tooley, Fry)	B453	A1	I a	g130				*St*	Toluca, Mexico (MEX-3)	1983	T	-	A	A	Y	G	G	T	G	A	G	R	G	Y	-	G	G	A	A
P9699	#15 (Camyon)	B354	A2	I a	g158				*St*	Sweden (SWE-2)	1998	C	G	A	W	C	G	T	Y	S	W	G	R	R	T	T	G	G	A	A
P10053	7MGK31/2 (Elansky)	B460	A1	II a	g154				*St*	Moscow Region, Russia (RUS-2)	1999	Y	G	R	A	Y	G	T	Y	S	T	*K*	A	R	C	T	G	-	A	A
P10105	US940507 (Fry)	A1129	A1	I b	US-1	US-1	86/100	92/100	*St*	Nebraska, USA (NE)	1994	*S*	*G*	G	W	Y	G	K	T	S	W	T	A	R	T	Y	R	G	M	A
P12001	1-40 (Fillipov)	A1318	A1	I a	–				*St*	Vologda Region, Russia (RUS-3)	2000	S	G	G	W	Y	G	K	T	S	W	T	A	R	T	*Y*	R	G	M	A
P8844	151 (Hohl, Pineda)	B99	A1	I b	US-1	US-1	86/100	92/100	*St*	Peru (PER)	1982	*S*	*G*	G	W	Y	-	K	T	S	W	T	A	R	T	Y	R	G	M	A
P10375	Phy2 (Yingfang)	B21	A1	I b	g144				*Le*	China (CHN)	~2003	*S*	*G*	G	W	Y	-	K	T	*S*	*W*	T	A	R	T	Y	R	G	M	A
P1375	130 (Hohl), ATCC60173	B464	A1	I b	g107				*St*	Switzerland (SUI)	~1982	*S*	*G*	G	W	Y	G	K	T	*S*	W	T	A	R	T	Y	-	G	M	A

Notes: [1] Mitochondrial haplotypes determined by method of Griffith & Shaw (1998)
[2] RAPD genotypes based on unpublished data (Coffey)
[3] DNA fingerprints based on published data using RG57 probe (Goodwin *et al.*, 1998)
[4] Isozyme patterns for glucose-6-phospate isomerase (*Gpi*) and peptidase (*Pep*)
**St* = *Solanum tuberosum, Le* = *Solanum lycopersicum.*
'–' Denotes missing data

confirmed that these isolates were genetically diverse by including representatives of both A1 and A2 mating types, all four mitochondrial haplotypes, a wide range of RAPD genotypes, and cultures with worldwide origins that were collected over several decades (M.D. Coffey, unpublished data) (Table 2). DNA was extracted as described previously (Blair *et al.*, 2008). PCR amplification and DNA sequencing of these isolates were done following the same protocols used during marker development, as already described.

Sequence data were generated for all isolates at each locus and were assembled into contigs using Sequencher 4.8 (Gene Codes). SNPs were identified by eye by the presence of heterozygous bases. For each locus, aligned sequences were numbered starting from either the forward PCR primer or the forward internal sequencing primer (Table 2), with two exceptions (3319_1R and 3408_8R): at these two loci, numbering began between the forward PCR primer and the internal sequencing primer. Only good quality, easily readable sequences were analysed and SNPs were scored using the 'Call Secondary Peaks' function in Sequencher 4.8 (Gene Codes). For SNP sites in regions where bidirectional sequence data were available, secondary peaks were called when their height was > 20% of the first peak in both directions. If a secondary peak was detectable but was smaller than this in either direction then that site in that isolate was excluded from analyses. For SNPs that occurred in regions where only one direction of sequence data were available, the secondary peak was scored only if its height was > 50% of the first peak and if the alternate/heterozygous base was found in at least one other isolate in the data set. More than 20 of the 32 isolates assessed needed to be scorable under these criteria for a SNP site to be identified. To minimize gaps in the data set used for population genetic analyses, 3.6% (n = 583) of SNPs scored in the 'straincode' (see below) across the 32 isolates were scored by eye as opposed to using the objective criteria just described.

Analysis of SNP data

Descriptive statistics calculated using the final data set from the international panel (Supplementary Table 1: online) included allele frequencies (calculated using TFPGA v1.3; Miller, 1997), number of diplotypes per locus, and nucleotide diversity, the latter two calculated using DnaSP 4.0 (Rozas *et al.*, 2003). Observed and expected heterozygosity were calculated manually for each SNP.

To minimize redundancy in information content associated with using linked SNPs, we explored the possibility of using only one SNP per locus for strain typing purposes. To maximize levels of polymorphism, we selected the SNP at each locus that had the highest frequency of the lesser allele; and to minimize any likelihood of physical linkage, we avoided using multiple MFR loci from the same super-contig of the whole genome sequence. Because using diploid data from linked SNPs presents analytical hurdles for population genetic analyses associated with unknown gametic phase data, we also used this data set of 19 unlinked SNPs opportunistically to explore the distribution of genetic diversity in the 32 isolate panel despite the fact that some ascertainment bias was introduced through the selection of the most polymorphic sites. Data from these 19 SNPs (Table 3 and Supplementary Table 1) were used for the following analyses as implemented in GenAlEx v6.1 (Peakall & Smouse, 2006): (i) tests for departures from Hardy–Weinberg equilibrium; (ii) a Mantel test to determine the statistical relationship between pairwise linear genotypic and geographic distance matrices; (iii) a principal coordinates analysis (PCO) on the squared genetic distance matrix; and (iv) probability of identity (*PI*) estimates, which assess the average probability that two unrelated individuals will share a genotype by chance using a given set of marker loci.

TFPGA v1.3 (Miller, 1997) was also used to calculate Rogers' original distance (Rogers, 1972). This was turned into a pairwise matrix of percentage similarity by subtracting the genetic distance from one (i.e. 1 – genetic distance). Per cent similarity between strains therefore represented the number of identical alleles between individual strains in a pair divided by the total number of alleles scored for both strains.

Applying SNP markers: Colombian isolates

As a preliminary look into how the SNP-based marker system might be used in an ecological context, we extracted DNA from 58 *P. infestans* isolates collected from six different *Solanum* species across a wide geographic range in Colombia. Colombia is geographically close to both areas that have been suggested as the centre of origin of the pathogen: Mexico's Toluca Valley and the South American Andes (Gómez-Alpizar *et al.*, 2007); hence we predicted that genetic diversity of isolates in Colombia might be high. We sequenced these isolates at the following five, SNP-rich MFRs, the first four of which contain sites used in the straincode: 3200F, 3230_1R, 2947R, 3345F and 3230_1F.

Table 3. Summary of data collected using the international panel of 32 isolates (see Table 2) for each MFR locus that generated useable data (i.e. including ones that did not contain SNPs, $n = 32$). Diversity levels at each locus are summarized by counts of the total number of SNPs (as judged by eye), the number of scored SNPs (as determined using the objective scoring criteria described in Materials and methods), and estimates of nucleotide diversity. The italicized number in parentheses given after nucleotide diversity reflects the number of isolates used in the calculation for that locus. Detailed primer information can be found in Table 1.

MFR Locus	Scored bp	Total SNPs	Scored SNPs	Nucleotide diversity	Diplo-types	GenBank Accession
2873_2R	505	0	0	0 (*23*)	–	HM166443 - HM166464
2908R	488	9	8	0.0024 (*32*)	4	HM166465 - HM166497
2947R	488	3	3	0.0021 (*32*)	4	HM166498 - HM166530
2995F	415	1	1	0.0009 (*32*)	2	HM166531 - HM166563
3010R	385	0	0	0 (*32*)	–	HM166564 - HM166595
3027R	501	4	4	0.0013 (*29*)	5	HM166596 - HM166628
3115F	483	2	2	0.0024 (*29*)	2	HM166629 - HM166661
3157F	497	5	3	0.0011 (*29*)	3	HM166662 - HM166694
3197F	394	2	1	0.0001 (*29*)	2	HM166695 - HM166725
3197R	444	3	2	0.00038 (*29*)	2	HM166726 - HM166755
3200F	454	21	16	0.013 (*30*)	6	HM166756 - HM166788
3200R	302	9	0	–	–	HM166789 - HM166820
3230_1F	430	11	9	0.0074 (*28*)	4	HM166821 - HM166852
3230_1R	446	11	8	0.0087 (*29*)	4	HM166853 - HM166884
3252F	433	7	5	0.0011 (*30*)	4	HM166885 - HM166917
3279F	426	5	4	0.003 (*29*)	4	HM166918 - HM166950
3318F	334	5	2	0.0022 (*27*)	3	HM166951 - HM166982
3318R	448	3	2	0.0008 (*31*)	4	HM166983 - HM167015
3319_1R	541	5	4	0.001 (*26*)	4	HM167016 - HM167048
3332F	401	9	7	0.0039 (*21*)	6	HM167049 - HM167081
3345F	352	5	4	0.0045 (*30*)	4	HM167082 - HM167115
3345R	304	4	0	–	–	HM167116 - HM167147
3350_4F	372	5	2	0.0012 (*25*)	3	HM167148 - HM167174
3350_9F	480	3	2	0.00092 (*30*)	3	HM167175 - HM167207
3350_9R	280	3	0	–	–	HM167208 - HM167240
3361_7R	415	5	2	0.0014 (*32*)	3	HM167241 - HM167273
3384R	512	0	0	0 (*32*)	–	HM167274 - HM167306
3390F	343	8	7	0.0041 (*29*)	8	HM167307 - HM167338
3408_3R	464	4	3	0.0018 (*23*)	3	HM167339 - HM167367
3430_8R	423	0	0	0 (*30*)	–	HM167368 - HM167398
3451_4F	423	1	1	0.00033 (*31*)	2	HM167399 - HM167430
3477_2F	423	14	0	–	–	HM166412 - HM166442

Results

Identification of microsatellites using MISA

MISA identified 738 microsatellite repeats within the *P. infestans* genome sequence, and Primer3 was able to design primers for 649 of these. Of the identified amplicons, 303 (~50%) should amplify a unique region, as judged by there being a single BLAST hit for the amplicon, and 268 contained no internal duplication, as judged by the occurrence of linearly ordered, non-overlapping high-scoring sequence pairs within the BLAST hit. About 25% of the microsatellite loci were present two to three times, 12% three to four times, and 7.5% five to six times, showing a typical Poisson distribution.

This initial list of 268 sequences was further refined by manual inspection. Longer repeat arrays were selected over shorter ones because there is a positive relationship between repeat length and mutation rate (Primmer & Ellegren, 1998); however, whether allele length is positively correlated to mutation rates in the flanking regions is not known. Dinucleotides and tetranucleotide repeats

are much rarer in exonic regions as compared with trinucleotide repeats (Toth *et al.*, 2000) so in the interest of isolating selectively neutral markers we limited the number of trinucleotide repeats analysed (to seven of 34 loci analysed; Table 1).

Initial SNP discovery

A preliminary search for SNPs in MFRs generated some positive results. Ten to twelve isolates from Eastern Canada were scored at each of four MFRs with an average of 376 bp analysed per locus (totalling 1427 bp). Sequences were identical across isolates for each locus and a total of eight heterozygous bases were found (data not shown).

Description of final molecular data set

We generated MFR sequence data from 28 unique, putatively unlinked loci (27 of the 28 loci were found on unique supercontigs of the *P. infestans* genome sequence) (Table 1). For six loci, the 'second' flanking region was targeted for sequencing after initial comparisons using the 'first' flanking region suggested a high number of SNPs. Thus the final molecular marker set consisted of 34 MFRs: one MFR for each of 22 microsatellite repeat loci (i.e. either immediately upstream or immediately downstream from the repeat) and two MFRs (i.e. both the upstream and downstream flanking regions) for each of six microsatellite repeat loci (Table 1). The first four digits in the name given to each locus refer to the number of the supercontig in which the microsatellite repeat was found.

DNA sequence data were obtained from the 34 MFRs for the 32-isolate SNP-discovery international panel (described in Materials and methods) and the whole-genome-sequenced T30-4 strain. High quality sequences that were predominantly bi-directional and easily alignable were collected for 32 of these 34 MFRs. PCR products averaged 525 bp in length, with scorable lengths ranging from 280 bp to 541 bp and averaging 426 bp (Table 3). Of the 32 MFRs that yielded analysable data, four had no detectable SNPs and four had no consistently scorable SNPs, leaving a total of 24 MFRs (across 21 microsatellites) with scorable SNP data.

Assessment of marker variability

A total of 13 624 base pairs of sequence consensus data from 32 MFRs (approximately 1M bp of individual sequence data) were generated and examined for the international panel of 32 *P. infestans* strains. Across the 24 MFR loci with scorable SNPs, a total of 167 SNPs were identified by eye and 102 were scorable under the objective criteria described in Materials and methods (Table 3), yielding a range of zero to 16 scorable SNPs per MFR (mean 3.2; Table 3). Nucleotide diversity ranged from a low of zero to a high of 0.013 (one change per 78 bp), with an average across 28 scored loci of 0.0023 (± 0.0029 SD; one change per 426 bp). Isolates with any missing data had to be excluded from nucleotide diversity calculations; hence if data were missing at one SNP, the isolate was removed from the analysis for all SNPs at that locus. This resulted in an average of 28.4 (± 3.3 SD) strains/locus being included, with a minimum of 20 isolates at one locus (Table 3). This loss of information and including only SNPs scored using objective scoring criteria makes this a marked underestimate of nucleotide diversity; the true measure is certainly higher.

All SNPs were biallelic except for one that was triallelic (Supplementary Table 1). For biallelic SNPs (n = 101), the frequency of the lesser allele averaged 0.16 (± 0.13 SD) and ranged from a minimum of 0.015 (a single heterozygote isolate at that site) to a maximum of 0.50 (all isolates heterozygous at that site; Supplementary Table 1). The number of sampled diplotypes per MFR locus ranged from two to eight and averaged 3.7 (± 1.5 SD; n = 24).

SNP data from all 24 MFR loci were able to distinguish each of the 32 isolates in the international panel. The pairwise per cent similarity between isolates varied from a high of 99.4% similar (a single allelic change from homozygous to heterozygous at one SNP across 88 scored SNPs) to a low of 58.5% similar (73 allelic changes at 54 of 88 scored SNPs). Average per cent similarity between isolates was 79.0%, which corresponds to about 37 allelic changes across the 88 scored SNPs.

Strain typing: development of a 'straincode'

As the 102 SNPs analysed here were easily able to discriminate among each isolate in the test panel, we selected a subset comprised of one SNP per locus (totalling 19 putatively unlinked SNPs) and explored its potential utility as a 'straincode' for typing purposes. SNPs in the straincode were selected based on allelic diversity and frequency of the lesser allele (described in full in Materials and methods), the latter of which also corresponded to the SNP with the highest heterozygosity for that locus in all but two cases (loci 3200F and 3230_1R; Table 3). SNPs in the straincode are ordered by decreasing information content as determined by per-locus probability of identity estimates (Table 2). This, incidentally, exactly matched the order that was initially determined based on allelic diversity and frequency of the lesser

allele, and therefore lends confidence to our method of selecting SNPs for the straincode in the first place. Genotype frequencies were consistent with Hardy–Weinberg equilibrium at 15 of the 19 SNPs in the straincode (Table 3).

The 19-SNP straincode is based on 18 biallelic SNPs, which have three diploid states, and one triallelic SNP, which has six diploid states, leading to over 2 billion (2 324 522 934) potential straincodes. The probability of two randomly chosen isolates having the same genotype when all 19 MFR loci are considered is very low ($PI = 3.5 \times 10^{-6}$). A probability of identity of approximately 1% is achieved when only the first four loci in the straincode are used ($PI = 1.6 \times 10^{-2}$); to reduce this to 0.1%, the first seven loci are needed ($PI = 1.1 \times 10^{-3}$).

The 19-SNP straincode was able to distinguish 25 of the 32 isolates in the international panel with pairwise percentage similarity between isolates ranging from 48% to 100%. The remaining seven isolates formed two groups: one with two identical strains, which were both collected from tomatoes in New York in two consecutive years and that share the same mating type, mtDNA haplotype, but not RAPD genotype; and one group with five identical isolates (Table 2), which were sampled from both tomato and potato plants. This group of five isolates showed very little variation in the larger, 102-SNP dataset, with only one to four SNPs varying among them. All were mating type A1, and two shared the same RAPD genotype (Table 2). These five isolates had a wide geographic distribution (Russia, Peru, China, Switzerland, and Nebraska) and differed from all other members of the international panel by a minimum of 36 allelic changes. Incidentally, as RAPDs are known for problematic reproducibility, it is noteworthy that different RAPD genotypes were found in both of the two groups of isolates with identical straincodes.

Based on data from the 19 SNPs in the straincode, a Mantel test showed no significant correlation between genetic and geographic distance matrices ($P = 0.21$), which is reflected by the general absence of clustering of isolates sampled from similar locations in the PCO analysis (Fig. 1). The PCO plot depicts high diversity and an overall lack of genetic structuring within the 32 isolate international panel; sampled isolates are generally scattered across the two-dimensional space with minimal clustering (Fig. 1). The only semblance of clustering of isolates according to geography is a loose at best cluster of five isolates sampled from the US (NY-1, NY-2, PA, FL, and TX; Fig. 1).

Colombian and Eastern Canadian isolates

Based on the sequencing of five MFR loci, the 57 isolates sampled from Colombia were virtually identical: the only exception was that single isolate that varied (A instead of G for strain P14006/1032) at a SNP not detected in the international panel (3230_1F at 204bp) and at which the other sampled isolates were homozygous for G (Supplementary Table 2: online). What can be formed as the start of the Colombian straincode is C-A-YG, showing similarity to P6576 (Netherlands) and P10250 (Hungary).

Similarly, two Eastern Canadian isolates used during marker development and optimization were identical at 46 MFRs screened during preliminary analyses (data not shown). Their straincode was unique among those found

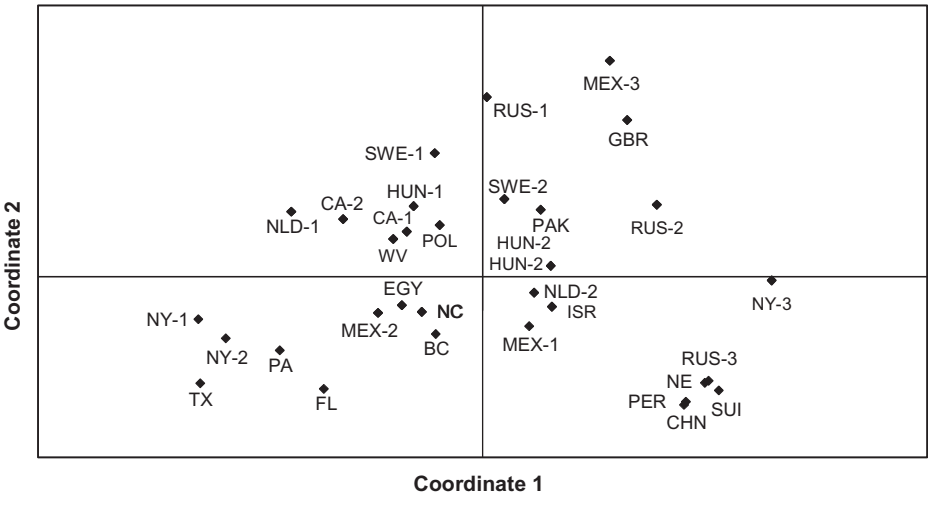

Fig. 1. Principal coordinates analysis plot of pairwise genotypic distance for 32 isolates of *Phytophthora infestans* scored at 19 unlinked SNPs in MFRs. Data labels reflect sampling location for each isolate as indicated in Table 2.

in the international panel (SGRWC-TTGWTGGTT-GGAA) and showed similarities to isolates P10157 (West Virginia) and P6747 (Poland). The dozen Canadian isolates that were used during the pilot development were also identical based on a smaller number of MFRs.

Discussion

SNP rates in P. infestans *MFRs*

Estimates of SNP rates are currently available for a small but growing number of taxa, though direct comparisons across studies are difficult when nucleotide diversity estimates are not provided. SNP rates are often expressed as the proportion of total base pairs surveyed that showed polymorphism (i.e. the number of variable sites divided by the number of bases analysed); but this is highly dependent on the sample size of the panel, thus comparisons across studies are inevitably confounded by variations in sample size. In contrast, because nucleotide diversity is the average number of different nucleotides per site between two randomly selected individuals in the panel, it is not dependent on sample size (apart from any effect of sampling error). Unfortunately, estimates of nucleotide diversity are absent more often than not from the relatively few published studies on SNP-based markers for ecological studies. For fungi and oomycetes, the following studies report the presence of SNPs but do not provide estimates of nucleotide diversity: Fisher *et al*. (1999) found 13 loci with one SNP each in the human pathogen, *Coccidioides immitis* G.W. Stiles, detected by single-strand conformational polymorphisms (SSCPs) in amplicons generated from random primers; Giresse *et al*. (2007) also used PCR-SSCP analysis of 124 ESTs from the sunflower pathogen, *Plasmopara halstedii* (Farl.) Berl. & De Toni, to find 25 SNPs across 12 loci; and Xu *et al*. (2007) found 178 SNPs in the basidiomycete *Tricholoma matsutake* (S. Ito & S. Imai) Singer through random sequencing of genomic library clones.

Overall nucleotide diversity of MFRs in *P. infestans* found here of 0.0023 (one change per 426 bp) is within the expected range of one SNP per 200–500 bp for non-coding DNA (Brumfield *et al*., 2003; Morin *et al.*, 2004) and is much higher than the broadly generalized rate of 1 SNP/1000 bp (Perkel, 2008). Tyler *et al*. (2006) found one change per 5000 bp in *P. ramorum* and one in 20 000 in *P. sojae* Kaufm. & Gerd. Comparing the annotated genome of *Pythium ultimum* Trow (Lévesque *et al.*, 2010) to another strain sequenced at low coverage by pyrosequencing, one change per 1121 bp was found (unpublished). A high rate of SNP was found in MFRs despite the fact that SNP rates reported here are considerable underestimates of true rates due to our use of strict objective scoring rules; 65 SNPs were detected by eye across the 24 polymorphic MFRs that were not scorable using the objective criteria. Further, isolates had to be removed from nucleotide diversity calculations if data were missing at even a single SNP. The loss of information content associated with objective scoring is clearly undesirable but needs to be balanced against the drawbacks of manual scoring, including human error and inconsistencies in judgement and that it may simply be impractical for large data sets when there is an automated alternative. Other estimates of nucleotide diversity in MFRs are limited to a single study on flycatchers: nucleotide diversity across six loci was around 0.005 (1 SNP every 200 bp) for the two species studied (Primmer *et al.*, 2002). To our knowledge, the only SNP-based markers in MFRs reported for a fungal species are three loci developed by Groenewald *et al.* (2007) for the plant pathogen *Cercospora beticola* Sacc. that were discovered incidentally while isolating size-polymorphic microsatellites. Although nucleotide diversity is not reported, a handful of other studies have developed SNP-based markers in MFRs and cover a diverse range of taxa, e.g. plants (Mogg *et al.*, 2002; Ablett *et al.*, 2006; Fernández *et al.*, 2008); birds (Primmer *et al.*, 2002); and abalone (Rhode *et al.*, 2008).

Regardless, it is reasonable to hypothesize that a likely contributor to the high SNP rate reported here is indeed that *P. infestans* regularly undergoes asexual reproduction and that there is no mechanism to purge new mutations from a clonal population. Indeed, human-mediated translocation of isolates and the bounty of hosts provided by agricultural monocultures give tremendous opportunity for rapid clonal propagation and accompanying genetic diversification through mutation. Whether the variation in the timing of collection of isolates used here contributed in any significant way to the high SNP rates is unknown, since potentially bottlenecked, clonal, pathogenic microbes are expected to have the capacity for rapid genetic change. However, that visual inspection of straincodes does not suggest the grouping of isolates by collection date and that the five isolates with the same straincode were collected from across at least 18 years suggest against a major effect of temporal variation in sample collection.

SNPs in MFRs for strain typing and population genetics

A range of molecular marker systems have been used for assessing genetic diversity in *P. infestans*, including allozymes (Goodwin *et al.*, 1995), RFLP fingerprinting (Goodwin *et al.*, 1992), AFLPs (Flier *et al.*, 2003), mtDNA haplotypes (Griffith & Shaw, 1998), RAPDs

(Mahuku *et al.*, 2000), and microsatellite repeat units (Lees *et al.*, 2006). The SNP-based straincode presented here offers the following advantages: (i) ease of scoring and reproducibility, (ii) low homoplasy, (iii) immediate compatibility with high throughput screening tools for SNPs (Kim & Misra, 2007; Beaudet & Belmont, 2008; Perkel, 2008), and (iv) the potential to characterize alleles directly from environmental samples such as spore traps. DNA sequencing is becoming the new standard to identify microorganisms. Similarly, an approach based on DNA sequences to identify strains of a given species would be more amenable to database search and unambiguous comparisons. An approach based on sequencing for large population studies might not be practically feasible now; however, next generation parallel sequencing technologies are developing rapidly and would allow direct evaluation of alleles present in samples from environmental metagenomics studies.

Results presented here suggest that SNPs in MFRs offer a promising avenue for strain typing. Probability of identity estimates indicate that only a small number (four) of the most informative SNP loci need to be screened to provide a unique genotype to 99 out of every 100 *P. infestans* isolates tested. However, it is important to bear in mind that probability of identity estimates are related to the diversity of the panel screened (see data below about the Colombian panel). If a strain typing system were to be developed for a more restricted geographic area (e.g. North America), its power to discriminate among isolates would need to be re-established through new probability of identity estimates using a panel of isolates from that region to avoid ascertainment bias affecting the results. This is the bias introduced to genetic analyses based on SNP data when arbitrary decisions are made during the SNP discovery phase; for example, if only the most variable SNP loci are analysed or if a non-representative sample of individuals is used when putative marker loci are screened for polymorphisms (Brumfield *et al.*, 2003).

The high genetic diversity of *P. infestans* at SNP markers described here suggests that they could be informative for answering molecular ecological questions about this pathogen, including population structure, the relative importance of sexual versus asexual reproduction, and questions associated with disease movements and relatedness among strains associated with different outbreaks. Indeed, in another oomycete plant pathogen, Delmotte *et al.* (2008) proved that an inordinate number of SNP loci is not necessarily required to successfully answer population genetics questions; they used 12 SNPs in ESTs to elucidate the relationship among pathotypes of *Plasmopara halstedii* (Farl.) Berl. & De Toni in France. Sampled pathotypes formed three genetic clusters that were consistent with a pattern of three historical introduction events followed by the subsequent emergence of new races, the evolution of which was hypothesized to have been mediated by both sexual (i.e. recombination) and asexual (i.e. accumulation of mutations) processes (Delmotte *et al.*, 2008). How many SNPs are needed to have an appropriate level of resolution? This is a question that is important to the scientific community in general (Smouse, 2010). Santure *et al.* (2010) found that 50 SNPs were as good as 20 microsatellite loci to estimate pedigree relatedness in a zebra fish population.

We found no evidence of overall genetic isolation by distance, which is not surprising given that human-mediated translocation is responsible for the geographic distribution of virtually the entire international panel. All isolates except one from Peru were sampled outside of the South American Andes region, which is now considered a possible region for the evolutionary origin of *P. infestans* (Gómez-Alpizar *et al.*, 2007). The most prominent genetic grouping of isolates is a cluster of five near-identical isolates sampled from Russia, Peru, China, USA and Switzerland that were collected up to 18 years apart: they differed from each other by a maximum of four nucleotide changes across 102 SNPs. It is reasonable to hypothesize that the similarity among these five isolates is because they are derived from the same original strain and that the small number of differences separating them occurred recently, after each translocated isolate became isolated and then spread through clonal propagation.

It is intuitive that our finding of no association between genetic similarity and geographic proximity would not hold for comparisons made at finer spatial scales. Indeed, all strains sampled from Colombia were genetically identical, thus they are likely all the same strain and products of clonal reproduction, which is supported by the fact that all isolates were mating type A1. As such, they do not have a genetic signature consistent with gene flow with the centre of origin. When this study was initiated, there was no knowledge of their genetic composition and it was a reasonable assumption that there might be considerable variation in genotypes. The very low level of variation observed here concurs with a results from a recent study focussed on *P. infestans* from Colombia, which found the same population to be highly clonal even with microsatellite analysis (Vargas *et al.*, 2009). However, the one strain that was different in our study did not match any of the two that were unique in Vargas *et al.* (2009) (Supplementary Table 1). Isolates from eastern Canada showed similar results with respect to lack of diversity on a regional scale: a limited data set

generated during marker development suggests that this group is also comprised of a single, clonal strain. These results are consistent with published studies indicating that populations of *P. infestans* in eastern Canada have been dominated by the US-8 genotype (A2 mating type) since its introduction into the region in the mid-1990s (Peters *et al.*, 1998, 1999, 2001). No A1 strains have been recovered from potatoes in this production area in over a decade (Peters *et al.*, unpublished data), and thus pathogen propagation is restricted to asexual reproduction and the spread of clones from initial disease foci generated from inoculum found in infected, over-wintered tubers used as seed or disposed of as culls.

This study shows that targeting sequence variation in MFRs can be an effective method of isolating SNP-based molecular markers. Sequence and SNP based approaches are likely to be the most common genotyping techniques in the future and an MFR SNP based system could provide a linkage and transition between the most common genotyping approach being currently used and this new generation of tools. That whole genome sequences are not available for most non-model organisms is becoming increasingly less limiting as new sequencing methods (Hudson, 2008) quickly transform our ability to cheaply acquire vast amounts of sequence data for these taxa. Similarly, many SNP genotyping technologies are now available and accommodate varying throughput levels (e.g. Behura, 2006; Kim & Misra, 2007; Beaudet & Belmont, 2008; Perkel, 2008; Brenan *et al.*, 2009; Drago *et al.*, 2009), which make screening large numbers of independent marker loci readily feasible even with environmental sampling such as spore traps. With SNP discovery and genotyping hurdles quickly vanishing, our ability to exploit the benefits of these abundant, low homoplasy, high throughput markers is continually enhanced.

Acknowledgements

The funding for this was provided by the Cbrne Research and Technology Initiative (project CRTI 04-0045RD). Grants to Michael Coffey from the USDA-NRI Plant Biosecurity program (2005-35605-15393 and 2008-55605-18773) were helpful in developing the Phytophthora DNA Bank which provided many samples for this project. We want to thank Carolyn Babcock, Ian Macdonald and Katie MacDonald for help with isolate culturing and Rafik Assabgui for the sequencing. We also want to thank Tyler Jackson and Hazim Alkhani for their technical assistance with the MFR sequencing of the isolates from eastern Canada and Colombia, respectively.

References

ABLETT, G., HILL, H., & HENRY, R.J. (2006). Sequence polymorphism discovery in wheat microsatellite flanking regions using pyrophosphate sequencing. *Mol. Breed., 17*, 281–289.

ADAMS, R.I., HALLEN, H.E., & PRINGLE, A. (2006). Using the incomplete genome of the ectomycorrhizal fungus *Amanita bisporigera* to identify molecular polymorphisms in the related *Amanita phalloides*. *Mol. Ecol. Notes, 6*, 218–220.

AITKEN, N., SMITH, S., SCHWARZ, C., & MORIN, P.A. (2004). Single nucleotide polymorphism (SNP) discovery in mammals: a targeted-gene approach. *Mol. Ecol., 13*, 1423–1431.

ALLAIN-BOULÉ, N., LÉVESQUE, C.A., MARTINEZ, C., BÉLANGER, R.R., & TWEDDELL, R.J. (2004). Identification of *Pythium* species associated with cavity-spot lesions on carrots in eastern Quebec. *Can. J. Plant. Pathol., 26*, 365–370.

BANKE, S., & MCDONALD, B.A. (2005). Migration patterns among global populations of the pathogenic fungus *Mycosphaerella graminicola*. *Mol. Ecol., 14*, 1881–1896.

BEAUDET, A.L., & BELMONT, J.W. (2008). Array-based DNA diagnostics: let the revolution begin. *Annu. Rev. Med., 59*, 113–129.

BEHURA, S.K. (2006). Molecular marker systems in insects: current trends and future avenues. *Mol. Ecol., 15*, 3087–3113.

BELFIORE, N.M., HOFFMAN, F.G., BAKER, R.J., & DEWOODY, J.A. (2003). The use of nuclear and mitochondrial single nucleotide polymorphisms to identify cryptic species. *Mol. Ecol., 12*, 2011–2017.

BENSCH, S., AKESSON, S., & IRWIN, D.E. (2002). The use of AFLP to find an informative SNP: genetic differences across a migratory divide in willow warblers. *Mol. Ecol., 11*, 2359–2366.

BLAIR, J.E., COFFEY, M.D., PARK, S.Y., GEISER, D.M., & KANG, S. (2008). A multi-locus phylogeny for *Phytophthora* utilizing markers derived from complete genome sequences. *Fungal Genet. Biol., 45*, 266–277.

BRENAN, C.J.H., ROBERTS, D., & HURLEY, J. (2009). Nanoliter high-throughput PCR for DNA and RNA profiling. *Methods Mol. Biol., 496*, 161–174.

BROHEDE, J., & ELLEGREN, H. (1999). Microsatellite evolution: polarity of substitutions within repeats and neutrality of flanking sequences. *Proc. R. Soc. Lond., Ser. B: Biol. Sci., 266*, 825–833.

BRUMFIELD, R.T., BEERLI, P., NICKERSON, D.A., & EDWARDS, S.V. (2003). The utility of single nucleotide polymorphisms in inferences of population history. *Trends Ecol. Evol., 18*, 249–256.

BURGESS, T., WINGFIELD, M.J., & WINGFIELD, B.W. (2001). Simple sequence repeat markers distinguish among morphotypes of *Sphaeropsis sapinea*. *Appl. Environ. Microbiol., 67*, 354–362.

COOKE, D.E.L., & LEES, A.K. (2004). Markers, old and new, for examining *Phytophthora infestans* diversity. *Plant Pathol., 53*, 692–704.

COOKE, D.E.L., LEES, A.K., HANSEN, J.G., LASSEN, P., ANDERSSON, B., & BAKONYI, J. (2009). From a European to a global database of *Phytophthora infestans* genetic diversity: examining the nature and significance of population change. *Acta Horticulturae, 834*, 19–26.

DELMOTTE, F., GIRESSE, X., RICHARD-CERVERA, S., M'BAYA, J., VEAR, F., TOURVIEILLE, J. Walser, P., & Tourvielle de Labrouhe, D. (2008). Single nucleotide polymorphisms reveal multiple introductions into France of *Plasmopara halstedii*, the plant pathogen causing sunflower downy mildew. *Infect., Genet. Evol., 8*, 534–540.

DOBROWOLSKI, M.P., TOMMERUP, I.C., SHEARER, B.L., & O'BRIEN, P.A. (2003). Three clonal lineages of *Phytophthora cinnamomi* in Australia revealed by microsatellites. *Phytopathology, 93*, 695–704.

DRAGO, F., KARPASITOU, K., & POLI, F. (2009). Microarray beads for identifying blood group single nucleotide polymorphisms. *Transfus. Med. Hemother., 36*, 157–160.

ELLEGREN, H. (2004). Microsatellites: simple sequences with complex evolution. *Nat. Rev. Genet., 5*, 435–445.

ENKERLI, J., WIDMER, F., GESSLER, C., & KELLER, S. (2001). Strain-specific microsatellite markers in the entomopathogenic fungus *Beauveria brongniartii*. *Mycol. Res., 105*, 1079–1087.

FERNÁNDEZ, M.P., NÚÑEZ, Y., PONZ, F., HERNÁIZ, S., GALLEGO, F.J., & IBÁÑEZ, J. (2008). Characterization of sequence polymorphisms from microsatellite flanking regions in *Vitis* spp. *Mol. Breed., 22*, 455–465.

FISHER, M.C., WHITE, T.J., & TAYLOR, J.W. (1999). Primers for genotyping single nucleotide polymorphisms and microsatellites in the pathogenic fungus *Coccidioides immitis*. *Mol. Ecol., 8*, 1082–1084.

FLIER, W.G., GRÜNWALD, N.J., KROON, L., STURBAUM, A.K., VAN DEN BOSCH, T.B.M., GARAY-SERRANO, E., . . . TURKENSTEEN, L.J. (2003). The population structure of *Phytophthora infestans* from the Toluca Valley of central Mexico suggests genetic differentiation between populations from cultivated potato and wild *Solanum* spp. *Phytopathology, 93*, 382–390.

FORBES, G.A., GOODWIN, S.B., DRENTH, A., OYARZUN, P., ORDONÌFEZ, M.E., & FRY, W.E. (1998). A global marker database for *Phytophthora infestans*. *Plant Dis., 82*, 811–818.

FRY, W.E., GOODWIN, S.B., MATUSZAK, J.M., SPIELMAN, L.J., MILGROOM, M.G., & DRENTH, A. (1992). Population genetics and intercontinental migrations of *Phytophthora infestans*. *Annu. Rev. Phytopathol., 30*, 107–129.

FRY, W.E., GOODWIN, S.B., DYER, A.T., MATUSZAK, J.M., DRENTH, A., TOOLEY, P.W., . . . SANDIAN, K.P. (1993). Historical and recent migrations of *Phytophthora infestans*: chronology, pathways, and implications. *Plant Dis., 77*, 653–661.

GARNICA, D.P., PINZÓN, A.M., QUESADA-OCAMPO, L.M., BERNAL, A.J., BARRETO, E., GRÜNWALD, N.J., & RESTREPO, S. (2006). Survey and analysis of microsatellites from transcript sequences in *Phytophthora* species: frequency, distribution, and potential as markers for the genus. *BMC Genomics, 7*, 245.

GIRESSE, X., TOURVIEILLE DE LABROUHE, D., RICHARD-CERVERA, S., & DELMOTTE, F. (2007). Twelve polymorphic expressed sequence tags-derived markers for *Plasmopara halstedii*, the causal agent of sunflower downy mildew. *Mol. Ecol. Notes, 7*, 1363–1365.

GLAUBITZ, J.C., RHODES, O.E., & DEWOODY, J.A. (2003). Prospects for inferring pairwise relationships with single nucleotide polymorphisms. *Mol. Ecol., 12*, 1039–1047.

GÓMEZ-ALPIZAR, L., CARBONE, I., & RISTAINO, J.B. (2007). An Andean origin of *Phytophthora infestans* inferred from mitochondrial and nuclear gene genealogies. *Proc. Natl. Acad. Sci. USA, 104*, 3306–3311.

GOODWIN, S.B., & DRENTH, A. (1997). Origin of the A2 mating type of *Phytophthora infestans* outside Mexico. *Phytopathology, 87*, 992–999.

GOODWIN, S.B., SCHNEIDER, R.E., & FRY, W.E. (1995). Use of celluloseacetate electrophoresis for rapid identification of allozyme genotypes of *Phytophthora infestans*. *Plant Dis., 79*, 1181–1185.

GOODWIN, S.B., SPIELMAN, L.J., MATUSZAK, J.M., BERGERON, S.N., & FRY, W.E. (1992). Clonal diversity and genetic differentiation of *Phytophthora infestans* populations in northern and central Mexico. *Phytopathology, 82*, 955–961.

GOODWIN, S.B., SMART, C.D., SANDROCK, R.W., DEAHL, K.L., PUNJA, Z.K., & FRY, W.E. (1998). Genetic change within populations of *Phytophthora infestans* in the United States and Canada during 1994 to 1996: role of migration and recombination. *Phytopathology, 88*, 939–949.

GRIFFITH, G.W., & SHAW, D.S. (1998). Polymorphisms in *Phytophthora infestans*: four mitochondrial haplotypes are detected after PCR amplification of DNA from pure cultures or from host lesions. *Appl. Environ. Microbiol., 64*, 4007–4014.

GRIMALDI, M.-C., & CROUAU-ROY, B. (1997). Microsatellite allelic homoplasy due to variable flanking sequences. *J. Mol. Evol., 44*, 336–340.

GROENEWALD, M., GROENEWALD, J.Z., LINDE, C.C., & CROUS, P.W. (2007). Development of polymorphic microsatellite and single nucleotide polymorphism markers for *Cercospora beticola* (Mycosphaerellaceae). *Mol. Ecol. Notes, 7*, 890–892.

GUO, J., VAN DER LEE, T., QU, D.Y., YAO, Y.Q., GONG, X.F., LIANG, D.L., . . . GOVERS, F. (2009). *Phytophthora infestans* isolates from Northern China show high virulence diversity but low genotypic diversity. *Plant Biol., 11*, 57–67.

HAAS, B.J., KAMOUN, S., ZODY, M.C., JIANG, R.H.Y., HANDSAKER, R.E., CANO, L.M., . . . NUSBAUM, C. (2009). Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature, 461*, 393–398

HENNEQUIN, C., THIERRY, A., RICHARD, G.F., LECOINTRE, G., NGUYEN, H.V., GAILLARDIN, C., & DUJON, B. (2001). Microsatellite typing as a new tool for identification of *Saccharomyces cerevisiae* strains. *J. Clin. Microbiol., 39*, 551–559.

HUDSON, M.E. (2008). Sequencing breakthroughs for genomic ecology and evolutionary biology. *Mol. Ecol. Resources, 8*, 3–17.

ISHIBASHI, Y., SAITOH, T., ABE, S., & YOSHIDA, M.C. (1996). Null microsatellite alleles due to nucleotide sequence variation in the grey-sided vole *Clethrionomys rufocanus*. *Mol. Ecol., 5*, 589–590.

IVORS, K., GARBELOTTO, M., VRIES, I.D.E., RUYTER-SPIRA, C., HEKKERT, B.T., ROSENZWEIG, N., & BONANTS, P. (2006). Microsatellite markers identify three lineages of *Phytophthora ramorum* in US nurseries, yet single lineages in US forest and European nursery populations. *Mol. Ecol., 15*, 1493–1505.

JONES, A.G., STOCKWELL, C.A., WALKER, D., & AVISE, J.C. (1998). The molecular basis of a microsatellite null allele from the white sands pupfish. *J. Hered., 89*, 339–342.

KIM, S., & MISRA, A. (2007). SNP genotyping: technologies and biomedical applications. *Annu. Rev. Biomed. Eng., 9*, 289–320.

KIMMEL, M., CHAKRABORTY, R., STIVERS, D.N., & DEKA, R. (1996). Dynamics of repeat polymorphisms under a forward-backward mutation model: within- and between-population variability at microsatellite loci. *Genetics, 143*, 549–555.

KNAPOVA, G., & GISI, U. (2002). Phenotypic and genotypic structure of *Phytophthora infestans* populations on potato and tomato in France and Switzerland. *Plant Pathol., 51*, 641–653.

LEE, S., & MOORMAN, G. (2008). Identification and characterization of simple sequence repeat markers for *Pythium aphanidermatum, P. cryptoirregulare*, and *P. irregulare* and the potential use in *Pythium* population genetics. *Curr. Genet., 53*, 81–93.

LEES, A.K., WATTIER, R., SHAW, D.S., SULLIVAN, L., WILLIAMS, N.A., & COOKE, D.E.L. (2006). Novel microsatellite markers for the analysis of *Phytophthora infestans* populations. *Plant Pathol., 55*, 311–319.

LÉVESQUE, C.A. (2007). Molecular diagnostics of soilborne fungal pathogens and applications to disease management. In Z.K. PUNJA, S. DEBOER, & H. SANFAÇON (Eds.), *Biotechnology and plant disease management* (pp. 146–164). Wallingford: CABI.

LÉVESQUE, C.A., BROUWER, H., CANO, L., HAMILTON, J.P., HOLT, C., HUITEMA, E., . . . BUELL, C.R. (2010). Genome sequence of the necrotrophic plant pathogen, *Pythium ultimum*, reveals original pathogenicity mechanisms and effector repertoire. *Genome Biology, 11*, R73.

LI, Y.C., KOROL, A.B., FAHIMA, T., BEILES, A., & NEVO, E. (2002). Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Mol. Ecol., 11*, 2453–2465.

LIEVENS, B., & THOMMA, B. (2005). Recent developments in pathogen detection arrays: implications for fungal plant pathogens and use in practice. *Phytopathology, 95*, 1374–1380.

MAHUKU, G., PETERS, R.D., PLATT, H.W., & DAAYF, F. (2000). Random amplified polymorphic DNA (RAPD) analysis of *Phytophthora infestans* isolates collected in Canada during 1994 to 1996. *Plant Pathol., 49*, 252–260.

MCCARTNEY, H.A., FOSTER, S.J., FRAAIJE, B.A., & WARD, E. (2003). Molecular diagnostics for fungal plant pathogens. *Pest Manage. Sci., 59*, 129–142.

MCEWEN, J.G., TAYLOR, J.W., CARTER, D., XU, J., FELIPE, M.S.S., VILGALYS, R., . . . SOARES, C.M.A. (2000). Molecular typing of pathogenic fungi. *Med. Mycol., 38*, 189–197.

MEGLECZ, E. (2007). MICROFAMILY (version 1): a computer program for detecting flanking-region similarities among different microsatellite loci. *Mol. Ecol. Notes, 7*, 18–20.

MEGLECZ, E., PETENIAN, F., DANCHIN, E., D'ACIER, A.C., RASPLUS, J.Y., & FAURE, E. (2004). High similarity between flanking regions of different microsatellites detected within each of two species of Lepidoptera: *Parnassius apollo* and *Euphydryas aurinia*. *Mol. Ecol., 13*, 1693–1700.

MHLANGA, M.M., & MALMBERG, L. (2001). Using molecular beacons to detect single-nucleotide polymorphisms with real-time PCR. *Methods, 25*, 463–471.

MILLER, M.P. (1997). *Tools For Population Genetic Analyses (TFPGA), version 1.3. A windows program for the analysis of allozyme and molecular population genetic data*. Flagstaff, AZ: Department of Biological Sciences, Northern Arizona University.

MOGG, R., BATLEY, J., HANLEY, S., EDWARDS, D., O'SULLIVAN, H., & EDWARDS, K.J. (2002). Characterization of the flanking regions of *Zea mays* microsatellites reveals a large number of useful sequence polymorphisms. *Theor. Appl. Genet.*, *105*, 532–543.

MOON, C.D., TAPPER, B.A., & SCOTT, B. (1999). Identification of epichloe endophytes in planta by a microsatellite-based PCR fingerprinting assay with automated analysis. *Appl. Environ. Microbiol.*, *65*, 1268–1279.

MORIN, P.A., LUIKART, G., & WAYNE, R.K. (2004). SNPs in ecology, evolution and conservation. *Trends Ecol. Evol.*, *19*, 208–216.

NARUM, S.R., BANKS, M., BEACHAM, T.D., BELLINGER, M.R., CAMPBELL, M.R., DEKONING, J., . . . GARZA, J.C. (2008). Differentiating salmon populations at broad and fine geographical scales with microsatellites and single nucleotide polymorphisms. *Mol. Ecol.*, *17*, 3464–3477.

PEAKALL, R., & SMOUSE, P.E. (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes*, *6*, 288–295.

PERKEL, J. (2008). SNP genotyping: six technologies that keyed a revolution. *Nat. Methods*, *5*, 447–453.

PETERS, R.D., PLATT, H.W., & HALL, R. (1998). Characterization of changes in populations of *Phytophthora infestans* in Canada using mating type and metalaxyl sensitivity markers. *Can. J. Plant. Pathol.*, *20*, 259–273.

PETERS, R.D., PLATT, H.W., & HALL, R. (1999). Use of allozyme markers to determine genotypes of *Phytophthora infestans* in Canada. *Can. J. Plant. Pathol.*, *21*, 144–153.

PETERS, R.D., FÖRSTER, H., PLATT, H.W.H., R., & COFFEY, M.D. (2001). Novel genotypes of *Phytophthora infestans* in Canada during 1994 and 1995. *Am. J. Potato Res.*, *78*, 39–45.

PRIMMER, C.R., & ELLEGREN, H. (1998). Patterns of molecular evolution in avian microsatellites. *Mol. Biol. Evol.*, *15*, 997–1008.

PRIMMER, C.R., BORGE, T., LINDELL, J., & SAETRE, G.P. (2002). Single-nucleotide polymorphism characterization in species with limited available sequence information: high nucleotide diversity revealed in the avian genome. *Mol. Ecol.*, *11*, 603–612.

RHODE, C., SLABBERT, R., & ROODT-WILDING, R. (2008). Microsatellite flanking regions: a SNP mine in South African abalone (*Haliotis midae*). *Anim. Genet.*, *39*, 329.

ROGERS, J.S. (1972). Measures of genetic similarity and genetic distance. *Studies in Genetics: Univ. Texas Publ. 7213*, *7*, 145–153.

ROZAS, J., SANCHEZ-DELBARRIO, J.C., MESSEGUER, X., & ROZAS, R. (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, *19*, 2496–2497.

ROZEN, S., & SKALETSKY, H. (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.*, *132*, 365–386.

SANTURE, A.W., STAPLEY, J., BALL, A.D., BIRKHEAD, T.R., BURKE, T., & SLATE, J. (2010). On the use of large marker panels to estimate inbreeding and relatedness: empirical and simulation studies of a pedigreed zebra finch population typed at 771 SNPs. *Mol. Ecol.*, *19*, 1439–1451.

SCHAAD, N.W., & FREDERICK, R.D. (2002). Real-time PCR and its application for rapid plant disease diagnostics. *Can. J. Plant. Pathol.*, *24*, 250–258.

SCHENA, L., CARDLE, L., & COOKE, D.E.L. (2008). Use of genome sequence data in the design and testing of SSR markers for *Phytophthora* species. *BMC Genomics*, *9*, 620.

SEDDON, J.M., PARKER, H.G., OSTRANDER, E.A., & ELLEGREN, H. (2005). SNPs in ecological and conservation studies: a test in the Scandinavian wolf population. *Mol. Ecol.*, *14*, 503–511.

SMITH, C.T., ELFSTROM, C.M., SEEB, L.W., & SEEB, J.E. (2005). Use of sequence data from rainbow trout and Atlantic salmon for SNP detection in Pacific salmon. *Mol. Ecol.*, *14*, 4193–4203.

SMITH, N.C., HENNESSY, J., & STEAD, D.E. (2001). Repetitive sequence-derived PCR profiling using the BOX-A1R primer for rapid identification of the plant pathogen *Clavibacter michiganensis* subspecies *sepedonicus*. *Eur. J. Plant Pathol.*, *107*, 739–748.

SMOUSE, P.E. (2010). How many SNPs are enough? *Mol. Ecol.*, *19*, 1265–1266.

STUKENBROCK, E.H., BANKE, S., & MCDONALD, B.A. (2006). Global migration patterns in the fungal wheat pathogen *Phaeosphaeria nodorum*. *Mol. Ecol.*, *15*, 2895–2904.

SULLIVAN, T.J., & FAETH, S.H. (2004). Gene flow in the endophyte *Neotyphodium* and implications for coevolution with *Festuca arizonica*. *Mol. Ecol.*, *13*, 649–656.

TERO, N., NEUMEIER, H., GUDAVALLI, R., & SCHLOTTERER, C. (2006). *Silene tatarica* microsatellites are frequently located in repetitive DNA. *J. Evol. Biol.*, *19*, 1612–1619.

TOTH, G., GASPARI, Z., & JURKA, J. (2000). Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res.*, *10*, 967–981.

TYLER, B.M., TRIPATHY, S., ZHANG, X., DEHAL, P., JIANG, R.H., AERTS, A. , . . . BOORE, J.L. (2006). *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. *Science*, *313*, 1261–1266.

VARGAS, A.M., QUESADA OCAMPO, L.M., CESPEDES, M.C., CARRENO, N., GONZALEZ, A., ROJAS, A., . . . RESTREPO, S. (2009). Characterization of *Phytophthora infestans* populations in Colombia: first report of the A2 mating type. *Phytopathology*, *99*, 82–88.

XU, J., GUO, H., & YANG, Z.L. (2007). Single nucleotide polymorphisms in the ectomycorrhizal mushroom *Tricholoma matsutake*. *Microbiology*, *153*, 2002–2012.

ZARDOYA, R., VOLLMER, D.M., CRADDOCK, C., STREELMAN, J.T., KARL, S., & MEYER, A. (1996). Evolutionary conservation of microsatellite flanking regions and their use in resolving the phylogeny of cichlid fishes (Pisces: Perciformes). *Proc. R. Soc. Lond., Ser. B: Biol. Sci.*, *263*, 1589–1598.

ZHOU, X.D., BURGESS, T.I., BEER, Z.W., LIEUTIER, F., YART, A., KLEPZIG, K., . . . WINGFIELD, M.J. (2007). High intercontinental migration rates and population admixture in the sapstain fungus *Ophiostoma ips*. *Mol. Ecol.*, *16*, 89–99.