

Article

Characterization of Plastidial and Nuclear SSR Markers for Understanding Invasion Histories and Genetic Diversity of *Schinus molle* L.

Rafael Plá Matielo Lemos ¹ , Cristiane Barbosa D'Oliveira Matielo ¹, Dalvan Carlos Beise ¹, Vanessa Gonçalves da Rosa ¹, Deise Schröder Sarzi ², Luiz Fernando Würdig Roesch ³ and Valdir Marcos Stefenon ^{1,*} 

- ¹ Núcleo de Ecologia Molecular e Micropropagação de Plantas, Universidade Federal do Pampa, São Gabriel ZIP 97307-020, Brazil; rafael.matielo@unipampa.edu.br (R.P.M.L.); crixdoliveira@gmail.com (C.B.D.M.); dalvanbio@gmail.com (D.C.B.); vaneessagr@hotmail.com (V.G.d.R.)
- ² Instituto de Bioquímica Médica Leopoldo de Meis—CCS, Universidade Federal do Rio de Janeiro, Rio de Janeiro 21941-902, Brazil; deisesarzi@hotmail.com
- ³ Centro Interdisciplinar de Pesquisas em Biotecnologia—CIP-Biotec, Campus São Gabriel, Universidade Federal do Pampa, São Gabriel ZIP 97307-020, Brazil; luizroesch@unipampa.edu.br
- * Correspondence: valdirstefenon@unipampa.edu.br; Tel.: +55-(55)-3237-0858

Received: 26 June 2018; Accepted: 25 July 2018; Published: 10 August 2018



Abstract: Invasive plant species are expected to display high dispersal capacity but low levels of genetic diversity due to the founder effect occurring at each invasion episode. Understanding the history of invasions and the levels of genetic diversity of such species is an important task for planning management and monitoring strategy for these events. Peruvian Peppertree (*Schinus molle* L.) is a pioneer tree species native from South America which was introduced in North America, Europe and Africa, becoming a threat to these non-native habitats. In this study, we report the discovery and characterization of 17 plastidial (ptSSR) and seven nuclear (nSSR) markers for *S. molle* based on low-coverage whole-genome sequencing data acquired through next-generation sequencing. The markers were tested in 56 individuals from two natural populations sampled in the Brazilian Caatinga and Pampa biomes. All loci are moderately to highly polymorphic and revealed to be suitable for genetic monitoring of new invasions, for understanding the history of old invasions, as well as for genetic studies of native populations in their natural occurrence range and of orchards established with commercial purposes.

Keywords: invasive species; genetic diversity; genetic resources; molecular markers

1. Introduction

The Peruvian Peppertree (*Schinus molle* L., Anacardiaceae; Figure 1) is a pioneer tree species native from South America and a threat to many environments around the world. This species can flower at any time due to its quick phenological cycle, which makes their seeds available for dispersal throughout the year [1,2]. Claimed to be an alien and invasive species, *S. molle* has plasticity and a natural ability to establish and survive in many environments. Well-known for its application in popular medicine [3], by the pharmaceutical uses of its essential oils [4], and by the production of the spicy pink pepper [5], *S. molle* has an important role in the medical, pharmaceutical and food industries. On the other hand, the reports of this species' invasion in California [6], Israel [7], Hawaii [8], South Africa [9], Mexico [10] and Italy [11] highlight the urgent need for tools to manage and control its dispersion capacity and large ability to establish in new environments. This scenario may become worse under the current climatic changes, as demonstrated by ecological niche modelling (ENM) of

S. molle distribution. A recent ENM study of *S. molle* potential distribution areas in Southern South America showed that there is a high probability of this species expanding over the Atlantic Forest [12]. Additionally, the risk of invasion of other of the world's biodiversity hotspots by this species is highly likely, due to its high adaptability into several environments.

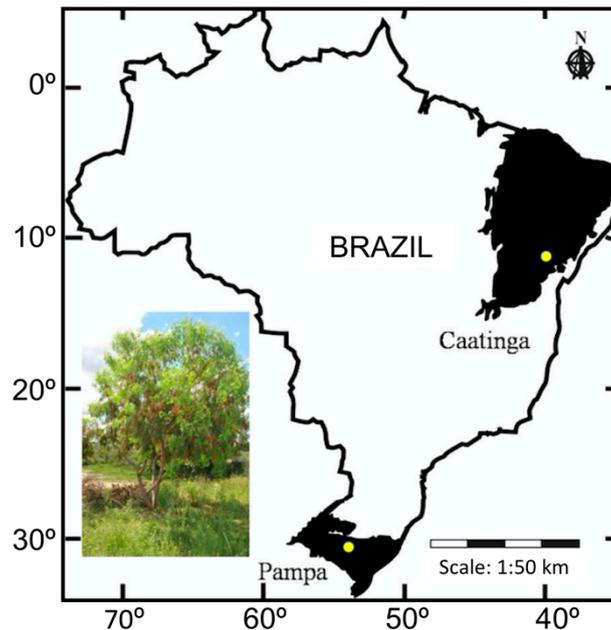


Figure 1. Location of the Pampa and Caatinga biomes in Brazil, where samples of *Schinus molle* were collected for this study. Dots within the biomes represent the location of each sampled population. Insert: fruiting adult tree of *S. molle*, within the Pampa biome.

Invasive species tend to present low genetic diversity due to the founder effect, i.e., the reduction in genetic diversity seen when a small subset of individuals splits from a larger and more diverse population, forming the basis of a new one [13]. Accordingly, low genetic diversity was observed in nine populations of *S. molle* studied in its natural range of occurrence in southern Brazil, based on dominant amplified fragments length polymorphism (AFLP) markers [14].

Mainly due to their locus-specific nature, co-dominance and high polymorphism, simple sequence repeats (SSR) markers have been used effectively for assessing genetic diversity, population differentiation, intra-population genetic structure, interpopulation gene flow, and fitness of natural populations (e.g., [15,16]). SSR markers can be characterized in both organellar or nuclear genomes. Organellar markers (plastidial and mitochondrial) have been widely used in phylogeographic studies, mostly due to the absence of recombination and the mainly uniparental (maternal) heritability. Neutral nuclear SSR markers (SSR loci found within non-expressed regions of the genome) are important to measure neutral DNA variation and are quite useful in populational studies.

There is a large number of SSR loci spread out all over the eukaryotes genomes, in coding and non-coding nuclear and organellar DNA [17]. Next-generation sequencing (NGS) platforms enable quick and cost-effective studies on small genomes, generating data of huge importance for biotechnological exploitation, conservation and breeding of non-model tree species, including the discovery of SSR markers [17]. Also, the use of molecular techniques has proved to be successful in reconstructing the history of invasions and evolutionary changes that may have occurred since the introduction of the species into non-natural environments [18].

Aiming to generate molecular markers for the close characterization of genetic diversity and structure, as well as the evolutive patterns and history of natural and invasive populations of *S. molle*, we employed NGS for the development of organellar and nuclear SSR markers. Here we report the

development and characterization of 24 SSR markers that presented moderate to high polymorphism when employed for genotyping adult individuals from two natural populations of *S. molle*. In addition, we discuss the outcomes of individual and combined use of the organellar (plastidial) and nuclear markers characterized.

2. Materials and Methods

2.1. Sample Strategies and DNA Extraction

Two natural populations of *Schinus molle* were mapped and collected in Pampa and Caatinga Biomes in Brazil (Figure 1). Twenty-seven individuals of population Pampa were sampled in the municipality of São Gabriel, Rio Grande do Sul State (30°21' S, 54°18' W), and 29 individuals were sampled in the Caatinga, Feira de Santana municipality, Bahia State (12°14' S, 38°57' W). Considering the commonly observed polymorphism of SSR loci and the literature related to characterization of SSR markers, this number of samples, equally distributed in two populations growing in different biomes enables the characterization of the polymorphism and usefulness of the discovered SSR loci at both, inter- and intrapopulation levels. Fresh leaves from each tree were sampled and individually conditioned in plastic bags with silica gel until the DNA extraction.

Total DNA was extracted using the protocol CTAB 2% [19] slightly modified. Quality of the isolated DNA was determined using a NanoVue™ spectrophotometer (GE Healthcare, Chicago, IL, USA). DNA was purified twice with the Agencourt® AMPure® XP Reagent (Beckman Coulter, Brea, CA, USA) and the final concentration of the DNA was quantified by using the Qubit Fluorometer kit (Invitrogen, Carlsbad, CA, USA) following the manufacturer's recommendations. The DNA samples were stored at −20 °C until use.

2.2. NGS Sequencing and de novo Assembly

A single plant of the population Pampa was selected for the NGS sequencing. An herbarium voucher of this sample was deposited in the Bruno Edgar Irgang Herbarium (HBEI) of the Federal University of Pampa under the number 1572. Total DNA from this plant was used for library preparation with Ion OneTouch™ 2 System using the Ion PGM™ Template OT2 400 Kit (Thermo Fisher Scientific, Waltham, MA, USA). The sequencing was performed using Ion PGM™ Sequencing 400 kit on the Ion PGM™ System with an Ion 318™ Chip v2.

The FastQ file exported from the Ion PGM™ System was evaluated for the K-mer number by KmerGenie software [20], and the *de novo* assembly was performed using the Velvet software [21].

2.3. Discovery and Characterization of SSR Markers

The software SSRLocator [22] was used to find di- and tri-nucleotide repeat motifs in the obtained contigs of the *S. molle* sample. The default parameters of SSRLocator were employed to identify SSR loci with a minimum of four and three repetitions for dimer and trimer motifs respectively. Primers for the identified SSR loci were designed using the Primer3 software [23], searching for alleles with sizes ranging from 90 to 500 bp. All contigs containing SSR loci were deposited in the GenBank (ID numbers are listed in Table 1).

Genomic origin of each SSR locus was identified through a BLAST search on the NCBI's website. For this search, each contig containing the SSR locus was compared to sequences of Viridiaeplantae deposited in the GenBank using the BLASTn tool and the MegaBLAST strategy. The bit-score and the percentage of identity of each contig to the subject of the BLAST analysis were used as conditions for determining the genomic origin of each SSR locus. According to the outcome of this search, SSR loci were classified as (i) a plastidial SSR locus (ptSSR) when the sequence was located within the plastome, or (ii) a nuclear neutral SSR locus (nSSR) when the contig failed to match with any sequence within the GenBank database. Such an outcome was interpreted as the contig being located in non-expressed regions within the nuclear genome.

Table 1. List of 24 SSR markers characterized for *Schinus molle*, including primer sequences (forward and reverse), repeat motif, length of the sequenced fragment, annealing temperature (Ta), genome region, GenBank accession number (GenBank ID), the bit-score and the identity of the contig to the subject sequence in the BLAST analysis.

Locus	Primer Sequence (5'→3') ^a	Rep Motif	Prod. Size	Ta (°C)	Genome Region ^b	GenBank ID	Bit-Score	Ident.
Smolle03	AAGTTTTATTTTCCCAGAAT AATCATAGGTTCTTCTCTCC	(TTC) ₃	169	51	ptSSR	MH536214	1050	99%
Smolle04	CTCCTAGGGATAAGAGACAT GAATAATTGTTGGAGACTCA	(AG) ₄	206	49	ptSSR	MH536215	941	99%
Smolle05	CGTAGACCAAATGATACAAT TTATTCTCATCAAACGAAT	(AGA) ₃	261	47	ptSSR	MH536216	941	99%
Smolle06	GGTCCATGAATCTAAGAAAT TTGAAATGAAATCTTTAGGA	(TC) ₄	140	47	ptSSR	MH536217	1014	100%
Smolle07	GAGTTGAAAATAAGCGTAGA TCTGGCTACTAAGATGTTTC	(CTG) ₃	117	51	ptSSR	MH536218	1022	99%
Smolle08	ATTTGTTATCTCATGTTTGC ACACATTGCTAACCAAATC	(GA) ₄	275	49	ptSSR	MH536219	970	99%
Smolle09	CCCATTAACATTTTAGAAGA GCTAAAAGTTGCAAAAATAAG	(GTT) ₃	188	47	ptSSR	MH536220	625	98%
Smolle10	TAGTTCATCCTATTGGCTC GAAACGAATTTTCATTTTAA	(TAA) ₃	234	47	ptSSR	MH536221	492	95%
Smolle11	AGAGGAGTAGTTATGAACCC TCACTATATTATCCITTTTCT	(GAA) ₃	140	51	ptSSR	MH536222	837	99%
Smolle12	CCACTAGAGATCAGAAAATG AATTGAGACGGTATTTTGTA	(AAG) ₃	128	47	ptSSR	MH536223	580	99%
Smolle13	CTGTGTTTTGGTAACAGTC GGTGGTAGGTAGAGAATAC	(CT) ₄	148	55	ptSSR	MH536224	1369	99%
Smolle14	AGTTTCTTTTACACATCCA AAGAAGATCCATTTGAGTT	(AT) ₄	161	47	ptSSR	MH536225	523	96%
Smolle15	GTACAAATAAGAATCCCCTT AGATCTGTAGCACTTACCA	(TTC) ₃	268	51	ptSSR	MH536226	1387	99%
Smolle16	GCAGATTCATCTAATTATGG AAGTTATAAGTTGTGAAGCG	(TCT) ₃	193	49	ptSSR	MH536227	588	98%
Smolle19	GTCAACTAAGGGGATAAGAT ATCCAATATCAATAAACCAA	(AT) ₆	159	47	ptSSR	MH536230	1061	97%
Smolle22	CTATAGTGGCTAGGGTGAG AATACCTTCTCTGTCTATCT	(AG) ₄	172	51	ptSSR	MH536232	483	94%
Smolle28	ATTTGTGCTCAATTTTCATA ACACATTGCTAACCAAATC	(GA) ₄	124	49	ptSSR	MH536237	494	96%
Smolle17	AGAATTCCTACCATTTCC ACAAAAATCACATGAAAATC	(TTC) ₃	203	47	nSSR	MH536228	-	-
Smolle18	TCTAGTACCAGAATCTTTGC AGGAGGTAATCCAATATC	(CT) ₅	127	51	nSSR	MH536229	-	-
Smolle21	GTGTTTCGTAAGACAAAAG GTGAGAAACGAATAAAGAAA	(TG) ₅	113	47	nSSR	MH536231	-	-
Smolle23	GAAGATAAGTTCATACCCT TTCAATTAATTGGCTCTAATC	(CTT) ₃	202	47	nSSR	MH536233	-	-
Smolle24	AGATTTCGGAATCTATTATT TGTTCAAGGAATAAAGGTAA	(TGA) ₃	226	47	nSSR	MH536234	-	-
Smolle25	TGCACCTTATATGAAAGACT ACCATCACTACAGCTCATA	(GAA) ₃	145	53	nSSR	MH536235	-	-
Smolle27	AGTCAATGAAGTTTTCACAG TGAGAACTCAAGATGCTATT	(CT) ₄	224	49	nSSR	MH536236	-	-

^a The forward primers contained the M-13 tail (5'-TGT AAA ACG ACG GCC AGT-3') at the 5' end.

^b ptSSR = plastidial markers; nSSR = nuclear neutral markers.

An *in silico* amplification was performed to test the potentially amplifiable SSR loci using the software SPCR [24]. With this strategy, we were able to identify primer pairs that were potentially amplifying a single locus in *S. molle* genome within the expected size range and discard primer pairs that were generating multi-loci amplifications and unfeasible band patterns. SSR loci presenting single locus amplifications in the *in silico* analyses were characterized in the 56 individuals from the sampled natural populations.

Selected SSR markers were amplified through PCR in a reaction mix with a final volume of 12.5 µL, which was prepared using an Eppendorf epMotion® 5070 (Eppendorf AG, Hamburg, Germany) containing about 50 ng of DNA, 0.25 µM of buffer, 0.5 µM of MgCl₂, 1 U of Taq DNA-Polymerase

(Invitrogen®), 0.05 µM of each dNTP, 0.125 µM of forward primer, 0.125 µM of reverse primer, and 0.125 µM of the M-13 primer (5'-TGTAACGACGGCCAGT-3') labelled with AlexaFluor 680 (Invitrogen®) fluorescence. The forward primer of each pair holds an M-13 tail, complementary to the fluorescent M13-primer (Table 1). Amplifications were carried out with an initial step of 94 °C for 3 min, a denaturing period of 94 °C for 45 s, annealing temperature ranging from 47 °C to 55 °C (see Table 1) for 30 s and extension at 72 °C for 1 min, for a total of 34 cycles, with a final extension step of 72 °C for 10 min in a BIO-RAD C1000 Touch™ Thermal Cycler (BioRad Co., Hercules, CA, USA).

SSR amplifications were electrophoretically separated on 6% polyacrylamide gels using a Li-Cor 4300 S automated DNA Sequencer® (LiCor Inc., Lincoln, NB, USA) and automatically scored and analyzed in SAGA^{GT} Software® (LiCor Inc.). An additional visual check of each scored gel was made in order to correct possible mismarked bands.

Total number of alleles (A), effective number of alleles (A_e), observed heterozygosity (H_O), expected heterozygosity (H_E), within population fixation index [$F_{IS} = (H_E - H_O)/H_E$], and deviation from Hardy-Weinberg equilibrium (HWE) were estimated for each locus in each population and overall. Differentiation between populations was estimated using the AMOVA approach. All estimations were performed using the software GenAEx 6.5 [25,26].

Linkage disequilibrium between pair of SSR loci was estimated using GenePop on web version 4.6 (<http://genepop.curtin.edu.au>) [27,28]. The Markov chain parameters used in the analysis were 1000 dememorization steps, 100 batches and 1000 iterations per batch.

3. Results

3.1. Sequencing Output and SSR Discovery

The low-coverage genome sequencing generated around 578 Mb of sequences that were used in the *de novo* assembly. A total of 1.267.774 bp with a GC content of 37% were obtained, corresponding to about 0.3X coverage of the *S. molle* genome, the estimated size of which is of about 410 Mbp [29]. Despite the low coverage of the genome, the sequencing enabled the discovery of more than 200 SSR sequences, 69 of which enabled the design of primers with a GC content of between 40 and 60%, closely matched annealing temperatures for forward and reverse primers, absence of undesirable secondary structures produced by intra- or intermolecular interactions (hairpins, self-dimer, cross-dimer) and generated a putative allele within the size range of 90 to 500 bp. These 69 loci were then further screened through *in silico* tests for amplification.

After the *in silico* test for amplification, 30 out of the 69 prospective SSR loci were considered optimal, presenting the amplification of a single locus within the expected range and were considered to be putative informative SSR markers. These 30 SSR loci were then characterized for polymorphism in two natural populations of *S. molle*.

3.2. Characterization of SSR Markers

Out of the 30 SSR markers tested in natural population samples, 24 presented amplifications within the expected range, while five markers failed at amplifying, even after replications of the PCR procedures. One SSR locus located within the mitochondrial genome of *S. molle* was excluded from further analysis. Of these 24 markers, 17 were located within the plastome (ptSSR markers), and seven in non-expressed regions within the nuclear genome (nSSR markers). The contigs containing the characterized SSR loci had a mean size of 540 bp and presented identity ranging from 94 to 100% of the contig (including the primers, the flanking regions and the SSR region) to the subject sequences of the BLAST analysis. All contigs presenting such high similarity to the subject sequences were located within the plastidial genome and were classified as plastidial SSR markers (ptSSR; Table 1). Seven contigs failed to match with sequences deposited in the GenBank. Because non-expressed nuclear intergenic regions are usually not deposited in this databank, we considered these seven SSR loci to be located within non-expressed regions of the nuclear genome of *S. molle* and classified them as neutral

nuclear SSR markers (nSSR; Table 1). The repeat motifs, forward and reverse primers, annealing temperature, percent of identity of the contigs to the subject sequence and the GenBank ID number of the primers of the 24 characterized SSR markers are listed in Table 1.

Moderate to high polymorphism was found in all 24 SSR markers tested in both populations (Table 2 and Supplementary Table S1). A total of 242 alleles were observed, ranging from 6 to 22 alleles per locus (Supplementary Table S1), with a mean A of 10.08 (Table 2); while the mean effective number of alleles per locus was $A_e = 6.03$, ranging from 2.52 to 10.56 (Supplementary Table S1). Expected and observed heterozygosities ranged from 0.60 to 0.90 (mean $H_E = 0.81$) and 0.00 to 1.00 (mean $H_O = 0.64$), respectively. The within-population fixation index (F_{IS}) ranged from -0.22 to 1.00 (mean $F_{IS} = 0.21$). Except for the loci Smolle09 and Smolle25 in the population Pampa, a significant deviation of HWE ($p < 0.05$) was observed in all loci in the overall analysis (Supplementary Table S1).

Table 2. Multilocus estimations of genetic diversity (A , A_e , H_O and H_E) and within population fixation index (F_{IS}). The mean number of samples (N) is given. Values are the average across 24 SSR loci.

	N	A	A_e	H_O	H_E	F_{IS}
Pampa	23.62	5.54	3.56	0.64	0.68	0.08
Caatinga	24.75	7.33	5.15	0.62	0.72	0.20
Overall	48.37	10.08	6.03	0.64	0.81	0.21

When evaluated individually in the population Caatinga, plastidial markers revealed lower estimations of genetic diversity (A , A_e , H_O and H_E) than neutral markers (Table 3). In the population Pampa, lower estimations were observed for the neutral nSSRs for H_O and H_E (Table 3). Congruently, the estimation of the within-population fixation index (F_{IS}) was higher for plastidial markers in the population Caatinga, while it was higher for nSSR markers in the population Pampa. When the different markers are combined, estimations of genetic diversity present low variation within each population (Table 3).

Table 3. Mean values of genetic parameters estimated for *Schinus molle* based on the 24 SSR markers characterized in this study. Data are presented independently for each genomic region, plastidial (ptSSRs) and neutral nuclear (nSSRs), for each population. Estimations include the number of samples (N), number of alleles per locus (A), effective allele number (A_e), observed (H_O) and expected (H_E) heterozygosities, and within-population fixation index (F_{IS}).

	ptSSRs		nSSRs	
	Caatinga	Pampa	Caatinga	Pampa
N	25.17	24.82	23.71	20.71
A	6.70	5.70	8.85	5.14
A_e	4.67	3.55	6.32	3.57
H_o	0.57	0.67	0.75	0.56
H_e	0.68	0.68	0.82	0.66
F_{IS}	0.25	0.03	0.10	0.22

The AMOVA approach revealed statistically significant ($p < 0.001$) differentiation between populations $F_{ST} = 0.13$ for all markers combined, $F_{ST} = 0.14$ for ptSSRs, $F_{ST} = 0.11$ for nSSRs, indicating a higher structure related to markers found in plastidial regions.

Estimations of linkage disequilibrium (LD) revealed 55 pairs of loci with significant ($p < 0.05$) LD in the population Pampa and 71 in the population Caatinga (Supplementary Table S2). Of these, 11 pairs of loci revealed significant LD in both populations (Table 4), suggesting physical ligation instead of inbreeding as the cause of this departure. Indeed, four out of the 11 pairs show ligation between plastidial markers and present physical ligation in the circular chromosome of the chloroplast.

However, seven pairs are formed by one plastidial and one nuclear marker, suggesting the effect of inbreeding, because physical ligation between plastidial and nuclear loci is not possible.

Table 4. Pairs of loci with statistically significant ($p < 0.05$) estimations of linkage disequilibrium (LD) in both populations. Supplementary Table S2 presents the list of all pairs of loci with significant LD. ptSSR markers are highlighted in red and nSSR markers in green.

Pairs of Loci		<i>p</i> -Values	
		Caatinga	Pampa
Smolle03	Smolle10	0.02184	0.02184
Smolle08	Smolle12	0.00356	0.00356
Smolle05	Smolle19	0.00111	0.00111
Smolle14	Smolle19	0.00000	0.00000
Smolle07	Smolle21	0.03328	0.03328
Smolle19	Smolle21	0.02626	0.02626
Smolle22	Smolle17	0.03454	0.03454
Smolle07	Smolle24	0.01180	0.01180
Smolle03	Smolle25	0.00057	0.00057
Smolle11	Smolle25	0.04062	0.04062
Smolle19	Smolle27	0.00000	0.00000

4. Discussion

In this study, we identified and characterized 17 ptSSR and seven nSSR markers for the invasive species *S. molle* based on low-coverage whole-genome sequencing. Non-model tree species usually lack information about genome sequences, making the development of molecular markers based on the knowledge of these data difficult and expensive. However, with the advance of NGS technologies in recent years, low-coverage whole-genome sequencing has been increasingly used for the development of SSR markers for such species. This approach was successfully used for developing SSR markers for several tree species with recognized economic importance [30–32] and, to our knowledge, this is the first report of using this approach for developing markers for an invasive tree species. Low coverage of the whole genome is a useful technique for sketching the genomic composition of a species [32], and even with particularly low-coverage sequencing, bioinformatics methodologies can generate extensive information about functional and repetitive elements, providing useful novel genomic resources [33].

The 24 SSR markers developed for *S. molle* presented moderate to high allelic diversity and heterozygosity in two natural populations, suggesting high usefulness for the characterization of genetic diversity and history in invaded areas. The two populations used in this study represent divergent biomes in Brazil and, therefore, a significant difference is expected between them, related to adaptation to the edafoclimatic conditions. The pairwise population differentiation estimated based on allele frequencies using the AMOVA approach was significant for each set of markers (ptSSRs, nSSRs) evaluated independently or combined. The slightly higher differentiation estimated with ptSSRs in comparison to nSSRs may be the effect of a selective sweep over some plastidial genes, affecting all SSR loci linked in the single plastidial chromosome in a genetic hitchhiking process. In fact, most of ptSSR loci revealed private alleles in both Pampa and Caatinga populations (Supplementary Table S3), supporting the existence of selection. Such private alleles may be very useful markers for determining the origin of individuals, and to reconstruct the history of invasion in recently occupied areas or even old invasions based on the frequency of particular alleles. These ptSSR markers may also be useful for differentiating populations growing under different selective pressure and to identify the origin of plants invading a determined environment.

In addition, assuming strict uniparental inheritance of the chloroplast in *S. molle*, the pollen-to-seed migration ratio (m_p/m_s) may be assessed from the plastidial and nuclear markers characterized in this study, using the estimations of F_{IS} obtained from nSSRs and F_{ST} obtained from each marker

class individually [34]. Using the F_{IS} and F_{ST} estimations from this study, m_p/m_s equaled 4.14 (see Supplementary file S4 for details about m_p/m_s computation), suggesting that seed dispersal plays a more important role than pollen in the total gene flow [34]. However, given the very large geographical distance splitting the Pampa and Caatinga populations, this estimation has no biological meaning, being used here merely for illustrative purposes.

The comparison between the two sets of markers did not diverge significantly in the patterns of diversity within each population. Recognizably low diversity and high fixation index were observed in population Caatinga, which experiences harder conditions in comparison to the population Pampa. Caatinga is a warm and dry biome with very specific and endemic xeric vegetal components [35,36], while the Pampa is a savanna-like environment with very few forest patches along its extension and a somewhat moist climate with four well characterized seasons [37].

Nuclear neutral SSRs are expected to present high levels of polymorphism due to their occurrence in the less conserved genomic regions. The ptSSR loci characterized for *S. molle* in this study presented basically the same estimations of nSSR loci, suggesting that despite the likely genetic hitchhiking process causing higher genetic differentiation at ptSSR markers, the selection is not strong when comparing the populations Caatinga and Pampa.

The elevated number of pairs of loci with significant LD observed in both populations is likely an effect of their isolation and small size, as also observed in the positive significant fixation index and the deviation from Hardy-Weinberg equilibrium in the ptSSR and nSSR markers for both populations. The linkage between pairs of ptSSR markers is not surprising, since they are located in the same circular plastidial chromosome. However, the significant LD for pairs of loci composed by plastidial and nuclear regions does not represent physical linkage, since they are located in different genomes.

5. Conclusions

Low-coverage whole-genome sequencing was employed in this study, with the aim of developing SSR markers for genetic studies of *S. molle*. The 24 SSR markers characterized here are moderately to highly variable and are useful for studies concerning the levels of diversity, the genetic differentiation, patterns of seed/pollen dispersal and the history of invasive populations of *S. molle* across the world, as well as for genetic studies of native populations in their natural occurrence range and of orchards established for commercial purposes.

Either employed independently or combined, both set of markers characterized in this study are very useful. Although linked in the single plastidial chromosome, ptSSR markers are highly informative and have been successfully used in phylogeographic studies, being the marker employed in 42% of the studies on this topic performed in the last decade [38], and also useful in population genetic diversity and structure analyses [39]. The neutral nuclear SSR markers, in turn, are quite useful in population genetic studies, since they are primarily not subjected to evolutionary forces and are likely to be distributed across the different chromosomes of *S. molle*, since no linkage disequilibrium between pairs of nSSR loci was observed.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2079-7737/7/3/43/s1>, Table S1: Genetic parameters estimated for *Schinus molle* based on the 24 SSR markers characterized in this study in overall populations and at population level; Table S2: Pairs of loci with statistically significant ($p < 0.05$) estimations of linkage disequilibrium (LD) in the populations Caatinga and Pampa; Table S3: Frequencies of private alleles in populations Pampa and Caatinga; File S4: Estimation of pollen-to-seed migration rate.

Author Contributions: Formal analysis, R.P.M.L., C.B.D.M., D.C.B. and V.G.d.R.; Funding acquisition, L.F.W.R. and V.M.S.; Investigation, R.P.M.L., D.S.S. and L.F.W.R.; Methodology, V.M.S.; Project administration, V.M.S.; Resources, R.P.M.L.; Supervision, V.M.S.; Writing—original draft, R.P.M.L.; Writing—review & editing, V.M.S.

Funding: This research was funded by Conselho Nacional de Desenvolvimento Científico e Tecnológico—CNPq/Brazil, grant number 442995/2014-8, Universidade Federal do Pampa—UNIPAMPA, Project number 09.011.16. The APC was funded by PROPP/UNIPAMPA.

Acknowledgments: The authors thank to João Vitor da Rosa by sampling individuals of the Caatinga population and CNPq/Brazil by the Research Fellowships provided to V.M.S. and L.F.W.R.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

- Milton, S.J.; Wilson, J.R.U.; Richardson, D.M.; Seymour, C.L.; Dean, W.R.J.; Iponga, D.M.; Procheş, Ş. Invasive Alien Plants Infiltrate Bird-Mediated Shrub Nucleation Processes in Arid Savanna. *J. Ecol.* **2007**, *95*, 648–661. [[CrossRef](#)]
- Iponga, D.M. Seed Set of the Invasive Tree *Schinus molle* (Anacardiaceae) in Semi-Arid Savanna, South Africa: The Role of Pollinators and Selfing. *J. Arid Environ.* **2010**, *74*, 414–416. [[CrossRef](#)]
- Dikshit, A.; Naqvi, A.; Husain, A. *Schinus molle*: A New Source of Natural Fungitoxicant. *Appl. Environ. Microbiol.* **1986**, *51*, 1085–1088. [[PubMed](#)]
- Marongiu, B.; Porcedda, A.P.S.; Casu, R.; Pierucci, P. Chemical Composition of the Oil and Supercritical CO₂ Extract of *Schinus molle* L. *Flavour Fragr. J.* **2004**, *19*, 554–558. [[CrossRef](#)]
- Goldstein, D.J.; Coleman, R.C. *Schinus molle* L. (Anacardiaceae) Chicha Production in the Central Andes. *Econ. Bot.* **2004**, *58*, 523–529. [[CrossRef](#)]
- Howard, L.F.; Minnich, R.A. The Introduction and Naturalization of *Schinus molle* (Pepper Tree) in Riverside, California. *Landsc. Urban Plan.* **1989**, *18*, 77–95. [[CrossRef](#)]
- Danin, A. The Inclusion of Adventive Plants in the Second Edition of Flora Palaestina. *Willdenowia* **2000**, *30*, 305–314. [[CrossRef](#)]
- Asner, G.P.; Jones, M.O.; Martin, R.E.; Knapp, D.E.; Hughes, R.F. Remote Sensing of Native and Invasive Species in Hawaiian Forests. *Remote Sens. Environ.* **2008**, *112*, 1912–1926. [[CrossRef](#)]
- Iponga, D.M.; Milton, S.J.; Richardson, D.M. Reproductive Potential and Seedling Establishment of the Invasive Alien Tree *Schinus molle* (Anacardiaceae) in South Africa. *Austral Ecol.* **2009**, *34*, 678–687. [[CrossRef](#)]
- Ramírez-Albores, J.E.; Bustamante, R.O.; Badano, E.I. Improved Predictions of the Geographic Distribution of Invasive Plants Using Climatic Niche Models. *PLoS ONE* **2016**, *11*, e0156029. [[CrossRef](#)] [[PubMed](#)]
- Stinca, A.; Chianese, G.; D’Auria, G.; Del Guacchio, E.; Fascetti, S.; Perrino, E.V.; Rosati, L.; Salerno, G.; Santangelo, A. New Alien Vascular Species for the Flora of Southern Italy. *Webbia* **2017**, *72*, 295–301. [[CrossRef](#)]
- Lemos, R.P.M.; D’Oliveira-Matielo, C.B.; Rodrigues, C.R.; Roesch, L.F.W.; Stefenon, V.M. Modeling distribution of *Schinus molle* L. in the Brazilian Pampa: Insights on vegetation dynamics and conservation of the biome. *Ann. For. Res.* **2014**, *57*, 205–214. [[CrossRef](#)]
- Ward, S.M.; Gaskin, J.F.; Wilson, L.M. Ecological Genetics of Plant Invasion: What Do We Know? *Invasive Plant Sci. Manag.* **2008**, *1*, 98–109. [[CrossRef](#)]
- Lemos, R.P.M.; D’Oliveira, C.B.; Stefenon, V.M. Genetic Structure and Internal Gene Flow in Populations of *Schinus molle* (Anacardiaceae) in the Brazilian Pampa. *Tree Genet. Genomes* **2015**, *11*, 75. [[CrossRef](#)]
- Nagel, J.C.; Ceconi, D.E.; Poletto, I.; Stefenon, V.M. Historical Gene Flow within and among Populations of *Luehea divaricata* in the Brazilian Pampa. *Genetica* **2015**, *143*, 317–329. [[CrossRef](#)] [[PubMed](#)]
- Stefenon, V.; Nagel, J.; Poletto, I. Evidences of Genetic Bottleneck and Fitness Decline in *Luehea divaricata* Populations from Southern Brazil. *Silva Fenn.* **2016**, *50*, 1566. [[CrossRef](#)]
- Taheri, S.; Lee Abdullah, T.; Yusop, M.; Hanafi, M.; Sahebi, M.; Azizi, P.; Shamshiri, R. Mining and Development of Novel SSR Markers Using Next Generation Sequencing (NGS) Data in Plants. *Molecules* **2018**, *23*, 399. [[CrossRef](#)] [[PubMed](#)]
- Meekins, J.F.; Ballard, J.H.E.; McCarthy, B.C. Genetic Variation and Molecular Biogeography of a North American Invasive Plant Species (*Alliaria petiolata*, Brassicaceae). *Int. J. Plant Sci.* **2001**, *162*, 161–169. [[CrossRef](#)]
- Doyle, J.; Doyle, J. A Rapid DNA Isolation Procedure for Small Quantities of Fresh Leaf Tissue. *Phytochem. Bull.* **1987**, *19*, 11–15.
- Chikhi, R.; Medvedev, P. Informed and Automated K-Mer Size Selection for Genome Assembly. *Bioinformatics* **2014**, *30*, 31–37. [[CrossRef](#)] [[PubMed](#)]
- Zerbino, D.R.; Birney, E. Velvet: Algorithms for de Novo Short Read Assembly Using de Bruijn Graphs. *Genome Res.* **2008**, *18*, 821–829. [[CrossRef](#)] [[PubMed](#)]

22. da Maia, L.C.; Palmieri, D.A.; de Souza, V.Q.; Kopp, M.M.; de Carvalho, F.I.F.; Costa de Oliveira, A. SSR Locator: Tool for Simple Sequence Repeat Discovery Integrated with Primer Design and PCR Simulation. *Int. J. Plant Genom.* **2008**, *2008*, 1–9. [[CrossRef](#)] [[PubMed](#)]
23. Untergasser, A.; Cutcutache, I.; Koressaar, T.; Ye, J.; Faircloth, B.C.; Remm, M.; Rozen, S.G. Primer3—New Capabilities and Interfaces. *Nucleic Acids Res.* **2012**, *40*, e115. [[CrossRef](#)] [[PubMed](#)]
24. Cao, Y.; Wang, L.; Xu, K.; Kou, C.; Zhang, Y.; Wei, G.; He, J.; Wang, Y.; Zhao, L. Information Theory-Based Algorithm for in Silico Prediction of PCR Products with Whole Genomic Sequences as Templates. *BMC Bioinform.* **2005**, *6*, 190. [[CrossRef](#)] [[PubMed](#)]
25. Peakall, R.; Smouse, P.E. Genalex 6: Genetic Analysis in Excel. Population Genetic Software for Teaching and Research. *Mol. Ecol. Notes* **2006**, *6*, 288–295. [[CrossRef](#)]
26. Peakall, R.; Smouse, P.E. GenAlEx 6.5: Genetic Analysis in Excel. Population Genetic Software for Teaching and Research—An Update. *Bioinform. Oxf. Engl.* **2012**, *28*, 2537–2539. [[CrossRef](#)] [[PubMed](#)]
27. Raymond, M.; Rousset, F. GENEPOP (Version 1.2): Population Genetics Software for Exact Tests and Ecumenicism. *J. Hered.* **1995**, *86*, 248–249. [[CrossRef](#)]
28. Rousset, F. Genepop'007: A Complete Re-Implementation of the Genepop Software for Windows and Linux. *Mol. Ecol. Resour.* **2008**, *8*, 103–106. [[CrossRef](#)] [[PubMed](#)]
29. Pivetta, M. Da quaresmeira ao jerivá: Tamanho do genoma de 100 árvores brasileiras varia até 20 vezes. *Pesquisa FAPESP* **2005**, *115*, 42–44.
30. Owusu, S.A.; Staton, M.; Jennings, T.N.; Schlarbaum, S.; Coggeshall, M.V.; Romero-Severson, J.; Carlson, J.E.; Gailing, O. Development of Genomic Microsatellites in *Gleditsia triacanthos* (Fabaceae) Using Illumina Sequencing1. *Appl. Plant Sci.* **2013**, *1*, 1300050. [[CrossRef](#)] [[PubMed](#)]
31. Wu, Y.; Zhang, R.; Staton, M.; Schlarbaum, S.E.; Coggeshall, M.V.; Romero-Severson, J.; Carlson, J.E.; Liang, H.; Xu, Y.; Drautz-Moses, D.I.; et al. Development of Genic and Genomic Microsatellites in *Gleditsia triacanthos* L. (Fabaceae) Using Illumina Sequencing. *Ann. For. Res.* **2017**. [[CrossRef](#)]
32. Staton, M.; Best, T.; Khodwekar, S.; Owusu, S.; Xu, T.; Xu, Y.; Jennings, T.; Cronn, R.; Arumuganathan, A.K.; Coggeshall, M.; et al. Preliminary Genomic Characterization of Ten Hardwood Tree Species from Multiplexed Low Coverage Whole Genome Sequencing. *PLoS ONE* **2015**, *10*, e0145031. [[CrossRef](#)] [[PubMed](#)]
33. Rasmusen, D.A.; Noor, M.A.F. What can you do with 0.1× genome coverage? A case study based on a genome survey of the scuttle fly *Megaselia scalaris* (Phoridae). *BMC Genom.* **2009**, *10*, 382. [[CrossRef](#)] [[PubMed](#)]
34. Petit, R.J.; Duminil, J.; Fineschi, S.; Hampe, A.; Salvini, D.; Vendramin, G.G. Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Mol. Ecol.* **2005**, *14*, 689–701. [[CrossRef](#)] [[PubMed](#)]
35. Franca-Rocha, W.; Silva, A.D.; Nolasco, M.C.; Lobão, J.; Britto, D.; Chaves, J.M.; Rocha, C.D. Levantamento Da Cobertura Vegetal e Do Uso Do Solo Do Bioma Caatinga. In Proceedings of the Ananis XIII Simpósio Brasileiro Sensoriamento Remoto INPE, Florianópolis, Brazil, 21–26 April 2007; pp. 2629–2636.
36. de Oliveira, G.; Araújo, M.B.; Rangel, T.F.; Alagador, D.; Diniz-Filho, J.A.F. Conserving the Brazilian Semiarid (Caatinga) Biome under Climate Change. *Biodivers. Conserv.* **2012**, *21*, 2913–2926. [[CrossRef](#)]
37. Roesch, L.F.W.; Vieira, F.C.B.; Pereira, V.A.; Schünemann, A.L.; Teixeira, I.F.; Senna, A.J.T.; Stefenon, V.M. The Brazilian Pampa: A Fragile Biome. *Diversity* **2009**, *1*, 182–190. [[CrossRef](#)]
38. Morris, A.B.; Shaw, J. Markers in time and space: A review of the last decade of plant phylogeographic approaches. *Mol. Ecol.* **2018**, *27*, 2317–2333. [[CrossRef](#)] [[PubMed](#)]
39. Mensous, M.; Van de Paer, C.; Manzi, S.; Bouchez, O.; Baâli-Cherif, D.; Besnard, G. Diversity and evolution of plastomes in Saharan mimosoids: potential use for phylogenetic and population genetic studies. *Tree Genet. Genomes* **2017**, *13*, 48. [[CrossRef](#)]

