Mycologia, 107(1), 2015, pp. 217–226. DOI: 10.3852/13-278
© 2015 by The Mycological Society of America, Lawrence, KS 66044-8897

Examining the efficacy of a genotyping-by-sequencing technique for population genetic analysis of the mushroom *Laccaria bicolor* and evaluating whether a reference genome is necessary to assess homology

Andrew W. Wilson[1]
Chicago Botanic Garden, Plant Conservation Science, 1000 Lake Cook Road, Glencoe, Illinois 60022

Norman J. Wickett
Chicago Botanic Garden, Plant Conservation Science, 1000 Lake Cook Road, Glencoe, Illinois 60022; Northwestern University, Program in Biological Sciences, 2205 Tech Drive O.T. Hogan Hall Room 2-144, Evanston, Illinois 60208

Paul Grabowski
University of Chicago, Ecology and Evolution, 1101 E 57th Street, Chicago, Illinois 60637

Jeremie Fant
Chicago Botanic Garden, Plant Conservation Science, 1000 Lake Cook Road, Glencoe, Illinois 60022

Justin Borevitz
Australian National University, Research School of Biology, Canberra, Australia; University of Chicago, Ecology and Evolution, 1101 E 57th Street, Chicago, Illinois 60637

Gregory M. Mueller
Chicago Botanic Garden, Plant Conservation Science, 1000 Lake Cook Road, Glencoe, Illinois 60022

Abstract: Given the diversity and ecological importance of Fungi, there is a lack of population genetic research on these organisms. The reason for this can be explained in part by their cryptic nature and difficulty in identifying genets. In addition the difficulty (relative to plants and animals) in developing molecular markers for fungal population genetics contributes to the lack of research in this area. This study examines the ability of restriction-site associated DNA (RAD) sequencing to generate SNPs in *Laccaria bicolor*. Eighteen samples of morphologically identified *L. bicolor* from the United States and Europe were selected for this project. The RAD sequencing method produced anywhere from 290 000 to more than 3 000 000 reads. Mapping these reads to the genome of *L. bicolor* resulted in 84 000–940 000 unique reads from individual samples. Results indicate that incorporation of non-*L. bicolor* taxa into the analysis resulted in a precipitous drop in shared loci among samples, suggests the potential of these methods to identify cryptic species. F-statistics were easily calculated, although an observable ''noise'' was detected when using the ''All Loci'' treatment versus filtering loci to those present in at least 50% of the individuals. The data were analyzed with tests of Hardy-Weinburg equilibrium, population genetic statistics (F_{IS} and F_{ST}), and population structure analysis using the program Structure. The results provide encouraging feedback regarding the potential utility of these methods and their data for population genetic analysis. We were unable to draw conclusions of life history of *L. bicolor* populations from this dataset, given the small sample size. The results of this study indicate the potential of these methods to address population genetics and general life history questions in the Agaricales. Further research is necessary to explore the specific application of these methods in the Agaricales or other fungal groups.

Key words: Agaricales, Agaricomycetes, Fungi, GBS, RAD, restriction-site associate DNA

Introduction

Fungi are involved in a vast array of ecological interactions. They are also one of the most cryptic groups of multicellular eukaryotic organisms. Determining the size and number of individuals in a population and the genetic relationships among populations is needed for understanding microevolutionary and ecological processes. Studying fungal populations is best accomplished with molecular methods because of challenges identifying individuals (genets) based on sporocarp demographic patterns. However, relative to plants and animals, population genetic studies in fungi are lacking. Only 8% of the population genetic studies appearing in *Molecular ecology* 2009–2010 involved fungi (Guichoux et al. 2011) and less than 7% of the population genetics papers published in *Molecular Ecology Notes* 2001–2005 focused on fungi (Dutech et al. 2007). Furthermore, a review of landscape genetic studies stated that only 0.5% of these studies involved fungi (Storfer et al. 2010). A review of population genetics of ectomycorrhizal fungi demonstrated that the accumulation of studies in this field has grown linearly 1990–2009 (Douhan et al. 2011), but fungal population studies still are relatively rare. By basis of

Submitted 31 Aug 2013; accepted for publication 4 Oct 2014.
[1] Corresponding author. E-mail: awwilson@purdue.edu

comparison, the proportion of fungal population genetic studies 1990–2009 made up only 1–2% of population genetic studies as a whole (evaluation using ISI Web of Science and the search terms ''fungi population genetics'' and ''population genetics'').

The paucity of research in fungal population genetics has left mycologists with many unanswered questions as to how ecological roles, life histories and evolutionary relationships shape fungal populations. Considering there are more than 1 500 000 estimated species of fungi and all of the ways these fungi contribute to the environment—from ecosystem services, to associations with foundation species—the underrepresentation of population genetic studies in fungi needs to be addressed. Several factors might contribute to the dearth of fungal population genetic research including complications of collecting sufficient samples of these cryptic organisms and challenge of developing population genetic markers for studying fungi.

A small number of studies have attempted to characterize the population genetic structure of some Agaricales (mushroom-forming fungi). Studies of the ECM fungus *Russula brevipes* identified distinct populations only over very long distances (> 1500 km) (Bergemann and Miller 2002) but was unable to detect population structure at shorter distances (< 1 km) (Bergemann et al. 2006), both studies relied on only three and six SSR (short sequence repeats, AKA microsatellites), respectively. A study of European *Laccaria amethystina* populations using a combination of SSRs, a mitochondrial marker, and direct amplification of length polymorphism (DALP) was not able to detect reproductive isolation among populations separated by distances > 400 km (Roy et al. 2008). Similarly Keirle et al. (2011) showed that Hawaiian populations of the saprobe fungus *Rhodocollybia laulaha* were not shaped by the geological history of the Hawaiian Islands, although this analysis was limited to a single genetic marker. The restricted number of markers used in these studies limited the resolution and potential for a clear assessment of gene flow or understanding evolutionary history within these groups.

Although SSRs are commonplace in population genetic studies of plants and animals, their usefulness in studying fungal populations appears to be limited. Dutech et al. (2007) enriched the SSR libraries for 17 species of fungi and concluded that SSRs are not only harder to isolate from fungi but that they exhibit a low polymorphism relative to angiosperms and animals.

In the past decade single nucleotide polymorphisms (SNPs) have become increasingly popular for use in population genetic studies for plants and animals (Guichoux et al. 2011, DeFaveri et al. 2013) but have yet to be used for population studies of Agaricales. SNPs are easily identified, are co-dominant and occur in such high numbers that in some cases they can be more efficient than SSRs because they do not require large sample sizes to characterize variation within a population nor do they need common controls across studies and time (Guichoux et al. 2011). However, as with SSRs, accumulating sufficient SNP data in fungi has been problematic due to laborious Sanger sequencing of limited loci (Keirle et al. 2011).

Restriction-site associated DNA (RAD) sequencing is a relatively new method using high-throughput sequencing technology that allows for identification of SNPs by sequencing many small fragments in highly variable regions of the genome accessed through restriction digest (Miller et al. 2007). The more closely related a group of organisms are, the more restriction loci they will have in common throughout their genome. Because of this principal, RAD sequencing has been demonstrated as an effective way to generate homologous loci from which SNPs can be identified for population genetic analysis. Elshire et al. (2011) developed a simplified RAD sequencing approach for genotyping by sequencing (GBS). This method recovered > 25 000 and 436 biallelic SNP markers from maize and barley, respectively. The ability of these techniques to identify a large number of SNP loci in a genome has the potential to overcome the limitations of traditional marker development in fungal population genetics (Davey and Blaxter 2011, Davey et al. 2011, Elshire et al. 2011, Morris et al. 2011, DeFaveri et al. 2013).

This study tests the effectiveness of RAD sequencing by with the GBS approach developed by Morris et al. (2011) to detect SNP markers in a small sampling of the ectomycorrhizal mushroom species *Laccaria bicolor*. This study also examines the efficacy of this method given taxonomic ambiguities related to the test species. Last, this study uses the availability of the *L. bicolor* genome to test whether a reference genome is necessary for analyzing population genetic structure with SNPs in Agaricales.

MATERIALS AND METHODS

Sample selection and phylogenetic evaluation of samples.— Eighteen samples, consisting of *Laccaria* species morphologically identified as *L. bicolor* (n = 7) and *L. trichodermophora* (n = 11), were chosen for GBS analysis (TABLE I) based on their morphological and geographic similarity. These two species have been reported to be closely related, representing northern and southern USA taxa of *L. bicolor* sensu lato (Mueller and Gardes 1991). Phylogenetic relationships among samples were determined with nuclear

TABLE I. GBS summary statistics for *Laccaria bicolor* samples

Sample No.	Sample ID	State/ country	Location	Phylogenetic position	No. of reads	[a]No. of unique mapping reads	No. of reads that map > 1	Percent mapping reads	Percent unique mapping reads
1	AWW537	IL	Johnson Co.	Ingroup	497 052	160 374	229 352	78	32
2	AWW538	IL	Johnson Co.	Ingroup	600 371	190 694	285 601	79	32
3	**AWW543[b]**	**IL**	**Johnson Co.**	**Outgroup**	**714 956**	**204 410**	**238 343**	**62**	**29**
4	AWW566	MI	Upper Peninsula	Ingroup	821 643	262 119	407 883	82	32
5	**AWW567**	**MI**	**Upper Peninsula**	**Outgroup**	**3 605 365**	**968 356**	**1 258 202**	**62**	**27**
6	GMM7591	Russia	Komarav Forest	Ingroup	2 170 743	809 821	957 341	81	37
7	GMM7595	France	Vallon d. Arcillac	Ingroup	240 385	75 421	115 395	79	31
8	GMM7712	TX	Newton Co.	Ingroup	1 565 187	508 231	762 694	81	32
9	GMM7733	TX	Tyler Co.	Ingroup	2 777 739	806 968	1 420 597	80	29
10	GMM7735	TX	Tyler Co.	Ingroup	1 913 068	574 907	952 009	80	30
11	**GMM7648**	**TX**	**Tyler Co.**	**Outgroup**	**2 805 911**	**359 891**	**455 644**	**29**	**13**
12	**GMM7698**	**TX**	**Newton Co.**	**Outgroup**	**1 867 648**	**412 158**	**318 259**	**39**	**22**
13	**GMM7703**	**TX**	**Newton Co.**	**Outgroup**	**2 204 898**	**516 575**	**362 365**	**40**	**23**
14	GMM7716	TX	Newton Co.	Ingroup	1 719 098	512 546	888 120	81	30
15	GMM7734	TX	Tyler Co.	Ingroup	290 022	84 926	147 104	80	29
16	KH-LA06-004	LA	—	Ingroup	2 041 959	482 268	885 654	67	24
17	KH-LA06-011	LA	—	Ingroup	1 968 307	616 130	974 982	81	31
18	KH-LA06-013	LA	—	Ingroup	1 594 716	496 205	796 485	81	31

[a] Unique mapping reads refers to the SNPs that map to only one location in the genome.
[b] Boldface indicates outgroup taxa.

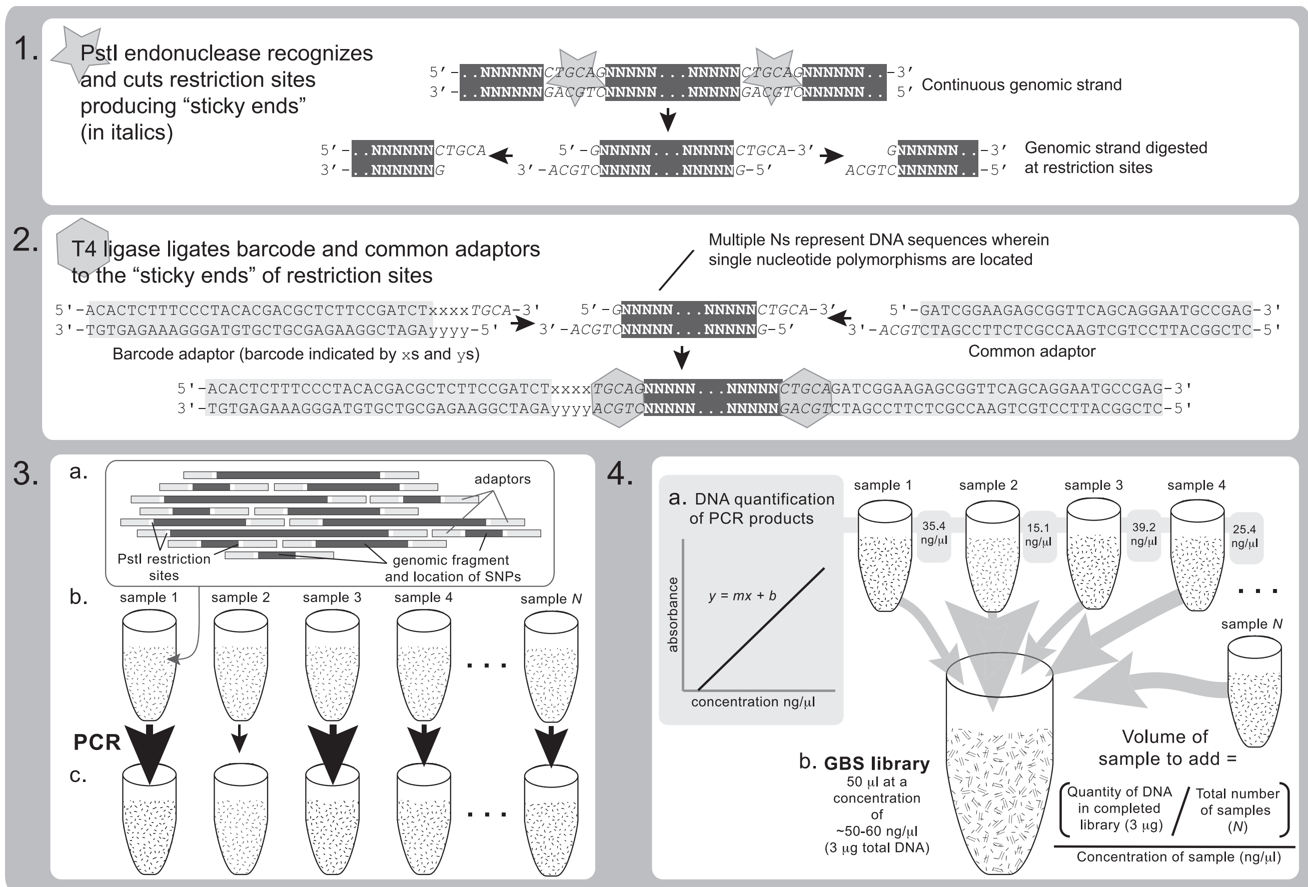

FIG. 1. GBS library preparation for high-throughput sequencing. Restriction digest step: The genome of each sample is digested using restriction endonuclease. For this study PstI is used. Dark gray regions with multiple, white, ''Ns'' denote genome sequence, which will be evaluated for viable SNPs.

FIG. 2. Ligation step: Barcode and common adapters are ligated to restriction sites. Adaptors are indicated by light gray shaded regions.

FIG. 3. PCR step: The ligated fragments are amplified before pooling into a single library. The arrows denote differences in successful amplification. Larger arrows = more amplified product.

FIG. 4. Quantification and library assembly step: The amount of DNA amplified in each sample is quantified with qPCR. The appropriate amount of sample then is added to a single 50 μL volume consisting of 3 μg DNA for sequencing. Arrows denote the volume of sample to be added to normalize the amount of each sample represented in the library. Larger arrows = greater volume of sample added. The sum of all samples should produce a final quantity of 3 μg DNA. This will be evaporated before being resuspended to produce a final concentration of 3 μg/50 μL DNA.

ribosomal internally transcribed spacer regions 1 and 2 (ITS) to determine their systematic and taxonomic relationship to 115 *Laccaria* and five outgroup samples (supporting materials TABLE I). Molecular methods for obtaining and analyzing nrITS sequence data follow the protocols outlined in Wilson et al. (2013).

GBS methods: restriction enzyme choice and adapter design.— The choice of restriction endonuclease (RE) used to digest genomic DNA depends on the size of the genome, the sequencing effort and the number of markers required for the necessary analyses. PstI, a 6 bp cutter (CTGCA|G), was chosen because it generates thousands of RAD loci in a fairly small genome such as *L. bicolor* (~ 65 megabases). This enzyme has been demonstrated to work well in switchgrass, *Panicum virgatum* (Morris et al. 2011).

Sequencing adapters consist of double-stranded oligonucleotides with 4 bp overhangs (or ''sticky ends'') that promote ligation to PstI restriction sites (FIG. 1) and allow for single- or paired-end, multiplex sequencing on the Illumina Inc. (San Diego, California) NGS platforms. These consist of a barcode adaptor and a common adaptor. Barcode adaptors terminate with a 4–8 bp barcode sequence adjacent to the 4 bp ''sticky end'' that corresponds to the PstI restriction sight (FIG. 2). The barcode adaptor sequences are: 5′–ACACTCTTTCCCTACACGACGCTCTTCCGATC-TxxxxTGCA and 5′– yyyyAGATCGGAAGAGCGTCGTG-TAGGGAAAGAGTGT where ''xxxx'' and ''yyyy'' correspond to the barcode and barcode reverse-complement sequences, respectively. The common adaptor simply has the PstI ''sticky end'' (in italics) with the sequence 5′–

CTCGGCATTCCTGCTGAACCGCTCTTCCGATCTGCA and a ''sticky end''-less reverse complement 5′–GATCG-GAAGAGCGGTTCAGCAGGAATGCCGAG. Adapter stocks consisting of both a barcode and the common adapter are made by combining both adapters, at a concentration of 0.6 ng/μL (~ 2 pmol/μL), mixed in a 1:1 ratio.

GBS method: library preparation and sequencing.—DNA quantification of samples is necessary to ensure that 1 μg genomic DNA is available for library preparation. For the library prep 10 μL ~ 100 ng/μL DNA is used per sample. Each sample of the appropriate DNA concentration is combined with 6 μL adapter stock described above (~ 0.06 pmol each adapter) and dried in 96-well plates for library preparation.

Four primary steps were involved in preparing a GBS library and follow the protocols described in Elshire et al. (2011), with some modification. These steps are illustrated (FIGS. 1–4) and are as follows: FIG. 1 Digesting *L. bicolor* genomic DNA in 20 μL reaction volumes using four units of Pst1-HF (NEB ref) per sample at 37 C for 2 h. FIG. 2. Ligating the restriction digest product of each sample with sequencing adapters in a 50 μL reaction volume using 640 units of T4 DNA ligase (NEB ref), incubating at 16 C for 60 min. FIG. 3. Purifying each sample with the QIAGEN MinElute 96-well PCR purification kit and protocol (QIAGEN USA, Valencia, California; http://www.qiagen.com), with the purified products eluted in 25 μL 0mM Tris-Cl, pH 8.5 (FIG. 3a, b). Following purification, PCR was performed on each sample in 50 μL volumes consisting of 25 μL NEB 2× Taq Master Mix, 2.0 μL PCR primer mix (12.5 pmol/μL of each of the following primers: (i) 5′–AATGATACGGCGACCACCGAGATCTACACTCTTTCCC-TACACGACGCTCTTCCGATC*T and (ii) –5′–CAAGCAGA-AGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAA-CCGCTCTTCCGATC*T, where * = phosphorothioation, 13 μL dH2O, and 10 μL purified post-ligation DNA samples (FIG. 3c). These thermo-cycler protocols were used for PCR amplification: (i) 68 C for 5 min; (ii) 95 C for 60 s; (iii) 95 C for 30 s; (iv) 65 C for 30 s; (v) 68 C for 30 s; (vi) repeat cycles 3–4 17 times; (vii) 68 C for 5 min; (viii) 4 C hold (FIG. 4). We quantified the concentration of PCR produce in each sample to ensure that each is represented equally when pooled into a singe library for multiplex sequencing. The concentration of each sample was calculated via qPCR with protocols developed for Quant-IT PicoGreen (Life Technologies, Grand Island, New York). The concentrations of each sample were used to determine the volume of sample to add to the final library. This was done with this formula:

$$V = (Q/N)/C$$

Where V = volume of sample to add to library (μL); Q = quantity of DNA in final library (3 μg); N = total number of samples; C = concentration of sample (ng/μL).

The pooled library was purified with a QIAQuick PCR purification kit (QIAGEN USA, Valencia, California; http://www.qiagen.com/), eluted with 50 μL buffer and quantified with a Nanodrop spectrophotometer (Thermo Scientific, Wilmington, Delaware). The library of *L. bicolor* samples was sequenced with an Illumina HiSeq 2000 at the Institute for Genomics and Systems Biology at the University of Chicago.

Analysis of RAD sequence data.—All data were deposited in NCBI's Bio Project archive No. PRJNA254545 (http://www.ncbi.nlm.nih.gov/bioproject/). Preliminary sequence data were evaluated with custom perl scripts. The software Stacks 1.02 (Catchen et al. 2011) was used for identification and quantification of SNP data. To evaluate the utility of the SNP data for population genetic studies of fungi other than *L. bicolor*, two treatments of the data were performed. The first treatment used the availability of the reference genome for *L. bicolor* (Martin et al. 2008). The other treatment estimated SNP data de novo to address the fact that most fungal species are not likely to have a readily available reference genome. Alignment to the reference genome was performed with Bowtie 1.0.0 (Langmead et al. 2009). To determine the utility of these data in population genetics studies, in particular the ability to calculate classical population genetics statistics, pairwise F_{ST} and F_{IS} scores were calculated for two North American metapopulations: Midwest (Illinois and Wisconsin) and South (Texas and Louisiana). To address missing data associated with the limited sampling of populations and individuals, we performed these analyses on two datasets: one using all loci (all loci dataset) and one using only those loci that are present in at least 50% of sampled individuals (50% dataset).

An evaluation of population structure from these SNP data was performed with Structure 2.3.4 (Pritchard et al. 2000), using the loci present in 50% dataset. Custom perl scripts were used to randomly subsample 1000 loci from both the de novo and reference-based treatments to directly compare results for equal numbers of loci. This was done 2× to compare the effect of subsampling loci. Structure was run with 20 000 MCMC generations with a burn-in of 10 000, using the methods described by Evanno et al. (2005) to test for the most likely population subdivision among k = 1–k = 7.

RESULTS

Selection of Laccaria bicolor *samples for RAD analysis.*—Thirteen of the 18 morphologically identified *L. bicolor* samples were shown to fall within the *L. bicolor* ''complex'' based on maximum likelihood analysis of nrITS sequences. This clade is given weak maximum likelihood bootstrap support (BS) in the nrITS tree (BS = 65%; SUPPLEMENTARY FIG. 1). However, this clade represents a diverse range of morphologically similar specimens, including the nrITS sequence for culture S238N-H82, which was used to produce the *L. bicolor* genome. The remaining five samples, originally identified as ''*L. bicolor*'', fall outside the complex. These represent three samples from Texas (GMM7648, GMM7698, GMM7730), one from Illinois (AWW543) and one from Michigan (AWW567). Despite the fact that these samples were misidentified as ''*L. bicolor*'' their placement outside

the *L. bicolor* complex is useful in that they were used in the evaluation of SNP data in intra- vs. interspecies relationships. OTUs from the monophyletic clade containing S238N-H82 were treated as the ingroup, while the five samples falling outside of the complex were treated as outgroups.

Evaluation of RAD data.—Initial assessment of RAD data (one multiplexed lane of paired-end HiSeq 2000) revealed that more than 64 000 000 101 bp sequence reads were produced from across all 18 samples. These sequences represent a total of 21 695 162 sites, from which preliminary analysis identified 17 854 SNPs (= 0.08% polymorphic sites). The percentage of sequence reads from ingroup *L. bicolor* complex samples that mapped to the genome was nearly double the percentage from outgroup taxa (ingroup average = 79.2%, outgroup mean = 46.4%: $P < 0.001$). When eliminating sequence reads that map with more than one genome location (e.g. using only reads with unique genome locations), the difference between ingroup and outgroup is still significant (ingroup average = 28.3%, outgroup mean = 22.8%: $P = 0.0025$). The number of informative SNP markers recovered varied depending on sample inclusion. Using both ingroup and outgroup samples reduces the number of informative SNP markers 94.4–98.2%, depending on how the population was circumscribed (data not included). To further evaluate the SNP markers generated by GBS methods, we focused on those produced for ingroup taxa.

All sequence reads were mapped to the genome for *L. bicolor* to determine the proportion of loci that fell within coding regions (TABLE II). Of all sequence reads, fewer than 8% were found in exons. Approximately one-quarter of the reads found in exons represent loci. These loci represent less than 20% of the loci used to evaluate population structure in this study.

A comparison of North American ingroup populations from the Midwest (two Illinois and one Michigan samples), and South (three each Texas and Louisiana samples) recovered 10 567 loci in the de novo treatment and 23 634 loci using a reference genome (TABLE I). Due to the small number of samples, this study cannot derive meaningful conclusions regarding the ecology and life history of *L. bicolor* populations in North America. Instead, the aim of the study was to use these data to demonstrate the kind of information obtainable from SNPs produced with GBS methods. All population genetics statistics generated for this dataset using STACKs, including expected and observed homozygosity and heterozygosity, Pi and the variance and standard error for these statistics etc.) are presented (SUPPLEMENTARY TABLE I).

The two datasets, all loci dataset and the 50% dataset, as applied to the two treatments, denovo and reference genome, were evaluated with population genetic statistics (TABLE II). In both datasets, more SNPs but fewer polymorphic loci were recovered using the reference genome compared to using the de novo method. Populations in the South (n = 8) had more than 4× as many loci as populations in the North (n = 3). Removing loci that were present in 50% or fewer individuals resulted in a 7× (de novo, Midwest) to 20× (reference, South) decrease in SNPs. This was as expected in that rare variants with low coverage can be identified readily with a reference genome but cannot be typed across many samples.

Fixation indices (F statistics) differed slightly between de novo and reference-based treatments in analyses that included all loci (TABLE II). The inbreeding co-efficient (F_{IS}) from de novo data for the Midwestern population is higher than the South (−0.0024 vs. −0.015) but both are effectively zero, suggesting little evidence of inbreeding. When the reference genome was used to filter data, the F_{IS} for Midwestern population showed a high heterozygous excess, but there was no difference seen for the Southern population (−0.370 vs. −0.0380). An excess of heterozygosity is usually associated with either hybridization, such as two divergent populations coming together, heavy selection against selfing associated with inbreeding depression or a product of small effective populations size, especially in dioecious species (Balloux 2004, Cabrera-Toledo et al. 2008). Because the excess of observed heterozygosity was found only in the Midwest this is more likely a data sampling issue rather than a potential indication of biological phenomena of populations in this area. Also, after removing loci not present in at least 50% of individuals, the difference between the observed and expected heterozygosity is reduced from approximately 100% more observed heterozygotes to roughly 33% more (TABLE II). Using loci that are present in only 50% of individuals or other such filtering of the datasets is suggested as an appropriate strategy to avoid confounding results (such as heterozygous excess) from the data. When applied to the dataset in this study, this approach produced much less variation between results derived from the de novo treatment or from the reference genome treatment.

Population fixation indices (F_{ST}) between de novo and reference treatments are not relatively different within datasets. However, there was a noticeable difference between datasets. Using all loci indicated that genetic structure among populations is high (F_{ST} = 0.225 vs. 0.217). In contrast, removing loci present in less than 50% of the individuals tended to limit

TABLE II. Results of population genetic analysis for SNP data from three North American *Laccaria bicolor* populations

	All loci (sequences with SNPs)				Loci in at least 50% of individuals			
	De novo		Reference		De novo		Reference	
US population	Midwest	South	Midwest	South	Midwest	South	Midwest	South
Total sites	2.15E+06	8.84E+06	5.80E+05	1.92E+06	6.35E+04	1.22E+05	2.24E+04	6.05E+04
Number of SNPs for the population	1680	9014	4083	19 811	212	952	209	958
Percent of total	0.078	0.102	0.704	1.033	0.334	0.783	0.934	1.583
Number of private alleles	570	824	548	768	11	39	17	56
Expected heterozygosity	0.326	0.401	0.417	0.454	0.325	0.276	0.317	0.306
Observed heterozygosity	0.632	0.649	0.825	0.853	0.456	0.287	0.481	0.400
F_{IS}	−2.40E-03	−1.50E-02	−3.70E-01	−3.80E-02	−2.60E-02	5.10E-02	−6.90E-02	−7.70E-02
Combined US population								
Number of SNPs	10 567		23 634		1140		1146	
Number of loci	6705		6287		704		383	
Pairwise F_{ST}	0.225		0.217		0.093		0.074	
Sequence reads that fall within exon regions and the proportion of those that consist of population genetic loci								
All sequence reads (with or without SNPs)	119 187		75 809					
Mapped to exons	4987		5626					
Percent of all reads	4.2%		7.4%					
Loci found in exons	1304		1029					
Percent of all loci	19.4%		16.4%					

extraneous ''noise'' from our dataset, clarified the results and suggested high gene flow between populations (F_{ST} = 0.093 vs. 0.074).

The Structure analyses were unable to find strong support for population structure using different genetic groupings for either the de novo or reference-based treatments, likely due to the small samples sizes (SUPPLEMENTARY FIG. 2). While there was some detectable variation in our populations according to Structure, interpretation of these results are limited by the amount of missing data in both the de novo treatment (63% missing) and reference genome treatment (65% missing). Despite the limitations of our data, it is likely that with a denser sampling strategy and appropriate filtering of the population genetic loci (e.g. the 50% dataset) the data generated with these GBS methods have the potential for using Structure on Agaricomycete population genetic data.

DISCUSSION

This is the first study demonstrating the efficacy of using RAD sequencing methods to develop SNP markers for population genetic studies in mushroom-forming fungi (Agaricales). Using the GBS method developed by Elshire et al. (2011), with modifications described in Morris et al. (2011), we generated thousands of viable RAD loci for a sample population of *Laccaria bicolor*, both with and without the use of a reference genome. These results demonstrate this method's potential use for analyzing Agaricales populations even when the target species lacks a genome reference.

As the cost of high-throughput sequencing continues to decline, it will become increasingly feasible to perform multiple sequencing runs on larger sample sizes both within and between populations. In addition, there is continued development of bioinformatic tools and programs appropriate for population genetic analysis of sequence data, which will make the pipelines for analysis of F statistics and other metrics increasingly straightforward. Efficient methods for generating population markers will facilitate and increase the use of these types of data for addressing big-picture questions, such as how life history and ecology play a role in shaping specific Agaricales populations.

Traditionally SSRs have been the marker of choice in population genetic studies, and they have been used effectively for some fungi (e.g. Bergemann and Miller 2002; Bergemann et al. 2006, 2009; Roy et al. 2008; Keirle et al. 2011; Vincenot et al. 2012). However, identifying informative SSRs for fungi has been shown to be difficult and time consuming (Dutech et al. 2007). The availability of high throughput sequencing of fungal genomes lets researchers identify SSRs by directly screening an organism's genome. For *L. bicolor* the availability of a genome facilitated the discovery of 78 SSRs of appropriate length from non-transposable element intergenic regions (Labbé et al. 2011). However, identification of SSRs is only the first of several steps required to develop SSRs as viable population genetic markers. Many individuals are needed to screen SSRs, assess allelic variation, ensure that they meet assumptions of Hardy-Weinberg equilibrium and detect the presence of null alleles. These are not trivial issues because many of these SSRs may not be variable across individuals (Dutech et al. 2007) or meet other required assumptions. We screened more than 20 potential SSRs from the *L. bicolor* genome, all of which were invariable across our tester strains (unpubl). In addition obtaining a sufficient number of samples can be problematic in cryptic organisms such as fungi. Despite the utility of SSRs in fungal population studies, their limitations in fungi and the limited progress in studying fungal populations cannot be overlooked. The development of new tools and analytical pipelines for measuring genetic variation within and among fungal populations will facilitate fungal population genetic studies and enable mycologists to explore questions of fungal life history and micro-evolutionary processes.

Research comparing the merits of SNP and SSR data (Rengmark et al. 2006, Coates et al. 2009, DeFaveri et al. 2013) have demonstrated the applications of both methods to the study of various organisms. Currently the utility of each marker is attributed to the scale in which the populations are to be evaluated. SSRs are favored for the evaluation of small-scale, intrapopulation structure evaluations of gene flow (Garvin et al. 2010, Hohenlohe et al. 2010, Davey et al. 2011) while SNPs are considered to be most effective at broad-scale, interpopulation and species-level evaluation (Narum et al. 2008, Hess et al. 2011, DeFaveri et al. 2013). Such comparison between SNP and SSR data has yet to be done for fungi.

This method also might help address the challenge of sufficiently sampling fungal populations. A recent study demonstrated that a high number of SNP loci generated with GBS can effectively measure genetic differentiation with F_{ST} values, even when sample sizes are small (Willing et al. 2012). As a result, GBS methods may benefit the study of fungal populations because the ability to generate hundreds to thousands of informative SNP loci will offset the inconsistent and ephemeral production of fungal sporocarps and other difficulties inherent in collecting fungal samples.

It is not clear how these methods might be used to study populations of asexual fungi. Genomic studies in asexual fungi suggest that SNP markers may be of limited value (Matteo Garbellotto pers comm). Although there is potential for large numbers of SNPs to identify somatic variation between genetically differentiated ''lines'' within a species, the quickly evolving nature of SSR markers potentially makes them more useful for addressing populations of asexually reproducing populations. Further evaluation of the GBS method is necessary to understand whether genomewide coverage of SNPs has value in measuring variation in asexually reproducing fungal populations.

GBS methods also may be useful for establishing the parameters around defining species and potentially can be used to understand the relationship within species complexes. Five of our 18 samples were identified to be outside the *L. bicolor* complex based on ITS sequence (SUPPLEMENTARY FIG. 1). Their inclusion in the analysis resulted in the attrition of around 95% of the RAD loci available from ingroup-only samples. Although restriction sites are somewhat conserved within species, they are much less so among species. Consequently the number of homologous restriction sites among species decreases significantly. The drop off of comparable restriction sites between unrelated taxa is useful in identifying cryptic species.

CONCLUSIONS

This study demonstrated the effectiveness of using RAD seq markers generated by the GBS method to identify SNPs in the mushroom species *Laccaria bicolor*. Thousands of loci based on SNPs were identified that can effectively measure intra- and interpopulation variation. In addition these methods demonstrated an appreciable drop off in common markers when evaluating interspecific relationships, meaning this method can be used to identify cryptic species and circumscribe species complexes. Currently SSRs are the most commonly used markers for evaluating fungal populations. However, because of documented challenges with developing informative SSRs and the need for large sample sets that may be difficult to obtain, the nearly exclusive dependency of using SSR markers for fungal population studies has impeded progress, and fungal population genetic research lags far behind that of animals and plants. The GBS method demonstrated here for generating RAD sequencing loci represents a new tool for evaluating fungal populations, and can help mycologists to better understand fungal life histories and micro-evolutionary processes.

ACKNOWLEDGMENTS

We wish to recognize the following: Nina Noah and Joe Davis for laboratory work and library prep of the samples; Matthew Johnson for data wrangling (AKA uploading sequence information to GenBank); Betty Strack for general research support.

LITERATURE CITED

Balloux F. 2004. Heterozygote excess in small populations and the heterozygote-excess effective population size. Evolution 58:1891–1900, doi:10.1111/j.0014-3820.2004.tb00477.x

Bergemann SE, Douhan GW, Garbelotto M, Miller SL. 2006. No evidence of population structure across three isolated subpopulations of *Russula brevipes* in an oak/pine woodland. New Phytol 170:177–184, doi:10.1111/j.1469-8137.2006.01654.x

———, Miller SL. 2002. Size, distribution and persistence of genets in local populations of the late-stage ectomycorrhizal basidiomycete, *Russula brevipes*. New Phytol 156: 313–320, doi:10.1046/j.1469-8137.2002.00507.x

———, Smith MA, Parrent JL, Gilbert GS, Garbelotto M. 2009. Genetic population structure and distribution of a fungal polypore, *Datronia caperata* (Polyporaceae), in mangrove forests of Central America. J Biogeography 36:266–279, doi:10.1111/j.1365-2699.2008.02006.x

Cabrera-Toledo D, Gonzales-Astorga J, Volvides AP. 2008. Heterozygote excess in ancient populations of the critically endangered Dioon caputoi (Zamiaceae, Cycadales) from central Mexico. Bot J Linn Soc 158:436–447, doi:10.1111/j.1095-8339.2008.00868.x

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. 2011. Stacks: building and genotyping loci de novo from short-read aequences. G3: Genes, genomes, genet 1:171–182.

Coates BS, Sumerford DV, Miller NJ, Kim KS, Sappington TW, Siegfried BD, Lewis LC. 2009. Comparative performance of single nucleotide polymorphism and microsatellite markers for population genetic analysis. J Hered 100:556–564, doi:10.1093/jhered/esp028

Davey JW, Blaxter ML. 2011. RADSeq: next-generation population genetics. Briefings in Funct Genomics 9: 416–423, doi:10.1093/bfgp/elq031

———, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. 2011. Genomewide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet 12:499–510, doi:10.1038/nrg3012

DeFaveri J, Viitaniemi H, Leder E, Merilä J. 2013. Characterizing genic and nongenic molecular markers: comparison of microsatellites and SNPs. Mol Ecol Res 13:377–392, doi:10.1111/1755-0998.12071

Douhan GW, Vincenot L, Gryta H, Selosse M-A. 2011. Population genetics of ectomycorrhizal fungi: from current knowledge to emerging directions. Fungal Biol 115:569–597, doi:10.1016/j.funbio.2011.03.005

Dutech C, Enjalbert J, Fournier E, Delmotte F, Barrés B, Carlier J, Tharreau D, Giraud T. 2007. Challenges of

microsatellite isolation in fungi. Fungal Genet Biol 44: 933–949, doi:10.1016/j.fgb.2007.05.003

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6:e19379, doi:10.1371/journal.pone.0019379

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. Mol Ecol 14:2611–2620, doi:10.1111/j.1365-294X.2005.02553.x

Garvin MR, Saitoh K, Churikov DY, Brykov VA, Gharrett AJ. 2010. Single nucleotide polymorphisms in chum salmon (Oncorhynchus keta) mitochondrial DNA derived from restriction site haplotype information. Genome 53:501–507, doi:10.1139/G10-026

Guichoux E, Lagache L, Wagner S, Chaumeil P, LÉGer P, Lepais O, Lepoittevin C, Malausa T, Revardel E, Salin F, Petit RJ. 2011. Current trends in microsatellite genotyping. Mol Ecol Res 11:591–611, doi:10.1111/j.1755-0998.2011.03014.x

Hess JE, Matala AP, Narum SR. 2011. Comparison of SNPs and microsatellites for fine-scale application of genetic stock identification of chinook salmon in the Columbia River Basin. Mol Ecol Res 11:137–149, doi:10.1111/j.1755-0998.2010.02958.x

Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genet 6:e1000862, doi:10.1371/journal.pgen.1000862

Keirle MR, Avis PG, Feldheim KA, Hemmes DE, Mueller GM. 2011. Investigation the allelic evolution of an imperfect microsatellite locus in the Hawaiian mushroom *Rhodocollybia laulaha*. J Hered 102:727–734, doi:10.1093/jhered/esr099

Labbé J, Murat C, Morin E, Le Tacon F, Martin F. 2011. Survey and analysis of simple sequence repeats in the *Laccaria bicolor* genome, with development of microsatellite markers. Curr Genet 57:75–88, doi:10.1007/s00294-010-0328-9

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25, doi:10.1186/gb-2009-10-3-r25

Martin F, Aerts A, Ahren D, Brun A, Danchin EGJ, Duchaussoy F, Gibon J, Kohler A, Lindquist E, Pereda V, Salamov A, Shapiro HJ, Wuyts J, Blaudez D, Buee M, Brokstein P, Canback B, Cohen D, Courty PE, Coutinho PM, Delaruelle C, Detter JC, Deveau A, DiFazio S, Duplessis S, Fraissinet-Tachet L, Lucic E, Frey-Klett P, Fourrey C, Feussner I, Gay G, Grimwood J, Hoegger PJ, Jain P, Kilaru S, Labbe J, Lin YC, Legue V, Le Tacon F, Marmeisse R, Melayah D, Montanini B, Muratet M, Nehls U, Niculita-Hirzel H, Secq MPO-L, Peter M, Quesneville H, Rajashekar B, Reich M, Rouhier N, Schmutz J, Yin T, Chalot M, Henrissat B, Kues U, Lucas S, Van de Peer Y, Podila GK, Polle A, Pukkila PJ, Richardson PM, Rouze P, Sanders IR, Stajich JE, Tunlid A, Tuskan G, Grigoriev IV. 2008. The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. Nature 452:88–92, doi:10.1038/nature06556

Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA. 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Res 17:240–248, doi:10.1101/gr.5681207

Morris GP, Grabowski PP, Borevitz JO. 2011. Genomic diversity in switchgrass (*Panicum virgatum*): from the continental scale to a dune landscape. Mol Ecol 20: 4938–4952, doi:10.1111/j.1365-294X.2011.05335.x

Mueller GM, Gardes M. 1991. Intra- and interspecific relations within the *Laccaria bicolor* sensu lato. Mycol Res 95:592–601, doi:10.1016/S0953-7562(09)80073-7

Narum SR, Banks M, Beacham TD, Bellinger MR, Campbell MR, Dekoning J, Elz A, Guthrie Iii CM, Kozfkay C, Miller KM, Moran P, Phillips R, Seeb LW, Smith CT, Warheit K, Young SF, Garza JC. 2008. Differentiating salmon populations at broad and fine geographical scales with microsatellites and single nucleotide polymorphisms. Mol Ecol 17:3464–3477.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

Rengmark AH, Slettan A, Skaala ϕ, Lie ϕ, Lingaas F. 2006. Genetic variability in wild and farmed Atlantic salmon (*Salmo salar*) strains estimated by SNP and microsatellites. Aquaculture 253:229–237, doi:10.1016/j.aquaculture.2005.09.022

Roy M, Dubois M-P, Proffit M, Vincenot L, Desmarais E, Selosse M-A. 2008. Evidence from population genetics that the ectomycorrhizal basidiomycete Laccaria amethystina is an actual multihost symbiont. Mol Ecol 17:2825–2838, doi:10.1111/j.1365-294X.2008.03790.x

Storfer A, Murphy MA, Spear SF, Holderegger R, Waits LP. 2010. Landscape genetics: Where are we now? Mol Ecol 19:3496–3514, doi:10.1111/j.1365-294X.2010.04691.x

Vincenot L, Nara K, Sthultz C, Labbé J, Dubois M-P, Tedersoo L, Martin F, Selosse M-A. 2012. Extensive gene flow over Europe and possible speciation over Eurasia in the ectomycorrhizal basidiomycete *Laccaria amethystina* complex. Mol Ecol 21:281–299, doi:10.1111/j.1365-294X.2011.05392.x

Willing E-M, Dreyer C, van Oosterhout C. 2012. Estimates of genetic differentiation measured by F_{ST} do not necessarily require large sample sizes when using many SNP markers. PLoS ONE 7:e42649, doi:10.1371/journal.pone.0042649

Wilson AW, Hosaka K, Perry BA, Mueller GM. 2013. *Laccaria* (Agaricomycetes, Basidiomycota) from Tibet (Xizang Autonomous Region, China). Mycoscience 54: 406–419, doi:10.1016/j.myc.2013.01.006