Published in final edited form as:

Mycologia. 2015 ; 107(1): 217–226. doi:10.3852/13-278.

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,

Chicago Botanic Garden, Plant Conservation Science, 1000 Lake Cook Rd, Glencoe, Illinois 60022, USA

Norman J. Wickett,

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

Paul Grabowski,

University of Chicago, Ecology and Evolution, 1101 E 57th Street, Chicago, IL 60637, USA

Jeremie Fant,

Chicago Botanic Garden, Plant Conservation Science, 1000 Lake Cook Rd, Glencoe, Illinois 60022, USA

Justin Borevitz, and

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

Gregory M. Mueller

Chicago Botanic Garden, Plant Conservation Science, 1000 Lake Cook Rd, Glencoe, Illinois 60022, USA

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 partially be explained by their cryptic nature and difficulty in identifying genets. Additionally, 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 K to over three million reads. Mapping these reads to the genome of *L. bicolor* resulted in 84 K to 940 K 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 using 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 data set, 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.

Keywords

restriction-site associate DNA; RAD; GBS; Fungi; Agaricomycetes; Agaricales

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 using 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* between 2009–2010 involved fungi (Guichoux et al. 2011), and less than 7% of the population genetics papers published in *Molecular Ecology Notes* between 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 recent review of population genetics of ectomycorrhizal fungi demonstrated that the accumulation of studies in this field has grown linearly between 1990–2009 (Douhan et al. 2011), fungal population studies are still relatively rare. By basis of comparison, the proportion of fungal population genetic studies between 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 1.5 million 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. There are several factors that may 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 (<1km) (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 saprobic 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 level of 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 (DeFaveri et al. 2013, Guichoux et al. 2011) but have yet to be used for population level 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 >25K 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 populations genetics (Davey and Blaxter 2011, Davey et al. 2011, DeFaveri et al. 2013, Elshire et al. 2011, Morris et al. 2011).

This study tests the effectiveness of RAD sequencing by using 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. Lastly, this study uses the availability of the *L. bicolor* genome to test whether a reference genome is necessary for analyzing population genetic structure using SNPs in Agaricales.

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 1) 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 using nuclear 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 1). 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 as it generates thousands of RAD loci in a fairly small genome like *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.I) and allow for single- or paired-end, multiplex sequencing on the Illumina Inc. (San Diego, CA) NGS platforms. These adaptors consist of a barcode adaptor and a common adaptor. Barcode adaptors terminate with a 4 to 8 bp barcode sequence adjacent to the 4bp “sticky end” that corresponds to the PstI restriction site. The barcode adaptor sequences are: 5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCTxxxxTGCA and 5'-yyyyAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT 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'-GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG. 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 together in a 1:1 ratio.

GBS method: library preparation and sequencing

DNA quantification of samples is necessary to ensure there is 1μg of genomic DNA available for library preparation. For the library prep, 10 μl of ~100 ng/μl DNA is used per

sample. Each sample of the appropriate DNA concentration is combined with 6 µl of an 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 in FIG. 1, I. – IV and are as follows: I. Digesting *L. bicolor* genomic DNA in 20 µl reaction volumes using 4 units of PstI-HF (NEB ref) per sample at 37 °C for 2 hours. II. 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 minutes. III. Purifying each sample using the Qiagen MinElute 96-well PCR purification kit and protocol (Qiagen USA, Valencia, CA. <http://www.qiagen.com>), with the purified products eluted in 25 µl 0mM Tris-Cl, pH 8.5. 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 A – 5' -

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG
ATC*T and B – 5' -

CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCT
TCCGATC*T, where “*” = phosphorothioation), 13 µl of dH₂O, and 10 µl of purified post-ligation DNA samples. The following thermocycler protocols were used for PCR amplification: 1) 68 °C for 5 minutes; 2) 95 °C for 60 seconds; 3) 95 °C for 30 seconds; 4) 65 °C for 30 seconds; 5) 68 °C for 30 seconds; 6) repeat cycles 3–4 17 times; 7) 68 °C for 5 minutes; 8) 4 °C hold. IV. Quantifying the concentration of PCR produce in each sample to ensure that each is represented equally when pooled into a single library for multiplex sequencing. The concentration of each sample was calculated via qPCR using protocols developed for Quant-IT PicoGreen (Life Technologies, Grand Island, NY). The concentrations of each sample were used to determine the volume of sample to add to the final library. This was done using the following formula:

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

Where, V = volume of sample to add to library (µl); Q = quantity of DNA in final library (3µg); 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, CA. <http://www.qiagen.com/>), eluted with 50µl elution buffer, and quantified with a Nanodrop spectrophotometer (Thermo Scientific, Wilmington, DE). The library of *L. bicolor* samples was sequenced using an Illumina HiSeq 2000 at the Institute for Genomics and Systems Biology at the University of Chicago.

Analysis of RAD sequence data

All data for this analysis have been deposited in NCBI's Sequence Read Archive (<http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?>) Bio Project ID number PRJNA254545. Preliminary sequence data were evaluated using custom perl scripts. The software Stacks v. 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 different 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 using a Bowtie v.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 meta-populations: 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 all sampled individuals (50% dataset).

An evaluation of population structure from these SNP data was performed with STRUCTURE v. 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 burnin of 10 000 and using the methods described by Evanno et al. (2005) to test for the most likely population subdivision among $k=1$ through $k=7$.

Results

Selection of *Laccaria bicolor* samples for RAD analysis

Thirteen of the 18 morphologically identified *L. bicolor* samples used in this study 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 of 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 of the *L. bicolor* complex is useful as they were used in the evaluation of SNP data in intra- vs. inter-species 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 over 64 million 101bp 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. only using 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 between 94.4% and 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 2). Of all sequence reads, less 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 (2 Illinois and 1 Michigan samples), and South (3 Texas and 3 Louisiana samples) recovered 10 567 loci in the de novo treatment, and 23 634 loci using a reference genome (TABLE 1). 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. Rather, the aim of the study was to use these data to demonstrate the kind of information obtainable from SNPs produced using GBS methods. All population genetics statistics generated for this data set using STACKs (including expected and observed homozygosity and heterozygosity, P_i , and the variance and standard error for these statistics, etc.) are presented in SUPPLEMENTARY TABLE 1.

The two datasets, all loci dataset and the 50% dataset, as applied to the two treatments, denovo and reference genome, were evaluated using population genetic statistics (TABLE 2). 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 $\times$ as many loci as populations in the North ($N=3$). Removing loci that were present in 50% of fewer of the individuals resulted in a 7 $\times$ (de novo, Midwest) to 20 $\times$ (reference, South) decrease in SNPs. This was as expected since rare variants with low coverage can be readily identified 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 2). 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 a hybridization event, 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). As 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 2). 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 extraneous “noise” from our dataset, clarified the results and suggested high levels of 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 using these GBS methods have the potential for using STRUCTURE on Agaricomycete population genetic data.

Discussion

This is the first study demonstrating the efficacy in 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 et al. 2006, Bergemann and Miller 2002, Bergemann et al. 2009, Keirle et al. 2011, Roy et al. 2008, 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 allows researchers to 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 to detect the presence of null alleles. These are not trivial issues since many of these SSRs may not be variable across individuals (Dutech et al. 2007) or meet other required assumptions. We screened over 20 potential SSRs from the *L. bicolor* genome, all of which were invariable across our tester strains (unpublished). Additionally, 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.

Previous research comparing the merits of SNP and SSR data (Coates et al. 2009, DeFaveri et al. 2013, Rengmark et al. 2006), 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, intra-population structure evaluations of gene-flow (Davey et al. 2011, Garvin et al. 2010, Hohenlohe et al. 2010) while SNPs are considered to be most effective at broad-scale, inter-population and species level evaluation (DeFaveri et al. 2013, Hess et al. 2011, Narum et al. 2008). Such comparison between SNP and SSR data has yet to be done for fungi.

This method may also help to 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 genome-wide coverage of SNPs has value in measuring variation in asexually reproducing fungal populations.

GBS methods may also be useful for establishing the parameters around defining species and can potentially be used to understand the relationship within species complexes. Five of our 18 samples were identified to be outside of 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 between species. Consequently, the number of homologous restriction sites between species decreases significantly. The drop off of comparable restriction sites between unrelated taxa is useful in identifying cryptic species.

Conclusion

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 inter-population variation. In addition, these methods demonstrated an appreciable drop off in common markers when evaluating inter-specific 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

- Balloux F. Heterozygote excess in small populations and the heterozygote-excess effective population size. *Evolution*. 2004; 58:1891–1900. [PubMed: 15521449]
- Bergemann SE, Douhan GW, Garbelotto M, Miller SL. No evidence of population structure across three isolated subpopulations of *Russula brevipes* in an oak/pine woodland. *New Phytologist*. 2006; 170:177–184. [PubMed: 16539614]
- Bergemann SE, Miller SL. Size, distribution, and persistence of genets in local populations of the late-stage ectomycorrhizal basidiomycete, *Russula brevipes*. *New Phytologist*. 2002; 156:313–320.
- Bergemann SE, Smith MA, Parrent JL, Gilbert GS, Garbelotto M. Genetic population structure and distribution of a fungal polypore, *Datronia caperata* (Polyporaceae), in mangrove forests of Central America. *Journal of Biogeography*. 2009; 36:266–279.
- Cabrera-Toledo D, Gonzales-Astorga J, Volvides AP. Heterozygote excess in ancient populations of the critically endangered *Dioon caputoi* (Zamiaceae, Cycadales) from central Mexico. *Botanical Journal of the Linnean Society*. 2008; 158:436–447.

- Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences. *G3: Genes, Genomes, Genetics*. 2011; 1:171–182. [PubMed: 22384329]
- Coates BS, Sumerford DV, Miller NJ, Kim KS, Sappington TW, Siegfried BD, Lewis LC. Comparative Performance of Single Nucleotide Polymorphism and Microsatellite Markers for Population Genetic Analysis. *Journal of Heredity*. 2009; 100:556–564. [PubMed: 19525239]
- Davey JW, Blaxter ML. RADSeq: next-generation population genetics. *Briefings in Functional Genomics*. 2011; 9:416–423. [PubMed: 21266344]
- Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*. 2011; 12:499–510.
- DeFaveri J, Viitaniemi H, Leder E, Merilä J. Characterizing genic and nongenic molecular markers: comparison of microsatellites and SNPs. *Molecular Ecology Resources*. 2013; 13:377–392. [PubMed: 23356957]
- Douhan GW, Vincenot L, Gryta H, Selosse M-A. Population genetics of ectomycorrhizal fungi: from current knowledge to emerging directions. *Fungal Biology*. 2011; 115:569–597. [PubMed: 21724164]
- Dutech C, Enjalbert J, Fournier E, Delmotte F, Barrés B, Carlier J, Tharreau D, Giraud T. Challenges of microsatellite isolation in fungi. *Fungal Genetics and Biology*. 2007; 44:933–949. [PubMed: 17659989]
- Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE*. 2011; 6:e19379. [PubMed: 21573248]
- Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*. 2005; 14:2611–2620. [PubMed: 15969739]
- Garvin MR, Saitoh K, Churikov DY, Brykov VA, Gharrett AJ. Single nucleotide polymorphisms in chum salmon (*Oncorhynchus keta*) mitochondrial DNA derived from restriction site haplotype information. *Genome*. 2010; 53:501–507. [PubMed: 20616872]
- Guichoux E, Lagache L, Wagner S, Chaumeil P, LÉger P, Lepais O, Lepoittevin C, Malausa T, Revardel E, Salin F, Petit RJ. Current trends in microsatellite genotyping. *Molecular Ecology Resources*. 2011; 11:591–611. [PubMed: 21565126]
- Hess JE, Matala AP, Narum SR. Comparison of SNPs and microsatellites for fine-scale application of genetic stock identification of Chinook salmon in the Columbia River Basin. *Molecular Ecology Resources*. 2011; 11:137–149. [PubMed: 21429170]
- Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags. *PLoS Genet*. 2010; 6:e1000862. [PubMed: 20195501]
- Keirle MR, Avis PG, Feldheim KA, Hemmes DE, Mueller GM. Investigation the Allelic Evolution of an Imperfect Microsatellite Locus in the Hawaiian Mushroom *Rhodocollybia laulaha*. *Journal of Heredity*. 2011; 102:727–734. [PubMed: 21914665]
- Labbé J, Murat C, Morin E, Le Tacon F, Martin F. Survey and analysis of simple sequence repeats in the *Laccaria bicolor* genome, with development of microsatellite markers. *Current Genetics*. 2011; 57:75–88. [PubMed: 21132299]
- Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009; 10:R25. [PubMed: 19261174]
- 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, Lucie E, Frey-Klett P, Fourrey C, Feussner I, Gay G, Grimwood J, Hoegger PJ, Jain P, Kilari 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. The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature*. 2008; 452:88–92. [PubMed: 18322534]
- Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research*. 2007; 17:240–248. [PubMed: 17189378]
- Morris GP, Grabowski PP, Borevitz JO. Genomic diversity in switchgrass (*Panicum virgatum*): from the continental scale to a dune landscape. *Molecular Ecology*. 2011; 20:4938–4952. [PubMed: 22060816]
- Mueller GM, Gardes M. Intra- and interspecific relations within the *Laccaria bicolor sensu lato*. *Mycological Research*. 1991; 95:592–601.
- Narum SR, Banks M, Beacham TD, Bellinger MR, Campbell MR, Dekoning J, Elz A, Guthrie CM Iii, Kozfkay C, Miller KM, Moran P, Phillips R, Seeb LW, Smith CT, Warheit K, Young SF, Garza JC. Differentiating salmon populations at broad and fine geographical scales with microsatellites and single nucleotide polymorphisms. *Molecular Ecology*. 2008; 17:3464–3477. [PubMed: 19160476]
- Pritchard JK, Stephens M, Donnelly P. Inference of Population Structure Using Multilocus Genotype Data. *Genetics*. 2000; 155:945–959. [PubMed: 10835412]
- Rengmark AH, Slettan A, Skaala Ø, Lie Ø, Lingaas F. Genetic variability in wild and farmed Atlantic salmon (*Salmo salar*) strains estimated by SNP and microsatellites. *Aquaculture*. 2006; 253:229–237.
- Roy M, Dubois M-P, Proffit M, Vincenot L, Desmarais E, Selosse M-A. Evidence from population genetics that the ectomycorrhizal basidiomycete *Laccaria amethystina* is an actual multihost symbiont. *Molecular Ecology*. 2008; 17:2825–2838. [PubMed: 18489549]
- Storfer A, Murphy MA, Spear SF, Holderegger R, Waits LP. Landscape genetics: where are we now? *Molecular Ecology*. 2010; 19:3496–3514. [PubMed: 20723061]
- Vincenot L, Nara K, Sthultz C, Labbé J, Dubois M-P, Tedersoo L, Martin F, Selosse M-A. Extensive gene flow over Europe and possible speciation over Eurasia in the ectomycorrhizal basidiomycete *Laccaria amethystina* complex. *Molecular Ecology*. 2012; 21:281–299. [PubMed: 22168318]
- Willing E-M, Dreyer C, van Oosterhout C. Estimates of Genetic Differentiation Measured by F_{ST} Do Not Necessarily Require Large Sample Sizes When Using Many SNP Markers. *PLoS ONE*. 2012; 7:e42649. [PubMed: 22905157]
- Wilson AW, Hosaka K, Perry BA, Mueller GM. *Laccaria* (Agaricomycetes, Basidiomycota) from Tibet (Xizang Autonomous Region, China). *Mycoscience*. 2013; 54:406–419.

FIGURE 1.

GBS library preparation for high-throughput sequencing. I. 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. II. Ligation step: Barcode and common adapters are ligated to restriction sites. Adaptors are indicated by light gray shaded regions. III. PCR step: The ligated fragments are amplified before pooling into a single library. The size of arrows denote differences in successful amplification. Larger arrows = more amplified product. IV. Quantification and library assembly step: The amount of DNA amplified in each sample is quantified using qPCR. The appropriate amount of sample is then added to a single 50 μl volume consisting of 3 μg of DNA for sequencing. The arrows denote the volume of sample to be added in order 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 of DNA. This will then be evaporated before being re-suspended to produce a final concentration of 3 μg/50 μl of DNA.

Table 1

GBS summary statistics for *Laccaria bicolor* samples. Unique mapping reads refers to the SNPs that map to only one location in the genome.

Sample Number	Sample ID	State/Country	Location	Phylogenetic Position	Number of Reads	Number of Unique Mapping Reads	Number of Reads that map > 1	% Mapping Reads	% 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	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

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			
	denovo		reference		denovo		reference	
	Midwest	South	Midwest	South	Midwest	South	Midwest	South
US population								
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	19811	212	952	209	958
% 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
number of SNPs	10567		23634		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)	119187	75809	
mapped to exons	4987	5626	
% of all reads	4.2%	7.4%	
loci found in exons	1304	1029	
% of all loci	19.4%	16.4%	