Patterns of genomic diversification reflect differences in life history and reproductive biology between figs (*Ficus*) and the stone oaks (*Lithocarpus*)

Journal:	<i>Genome</i>
Manuscript ID	gen-2016-0188.R1
Manuscript Type:	Article
Date Submitted by the Author:	09-Mar-2017
Complete List of Authors:	Kua, Chai-Shian; Xishuangbanna Tropical Botanical Garden, Key Lab in Tropical Ecology; Current address: The Morton Arboretum, Department of Science and Conservation Cannon, Charles; Xishuangbanna Tropical Botanical Garden, Key Lab of Tropical Ecology; Current address: The Morton Arboretum, Center for Tree Science
Is the invited manuscript for consideration in a Special Issue? :	Evolution of Tree Diversity
Keyword:	Tropical Biodiversity, Reference-free, comparative genomics, kmers, diversification

SCHOLARONE™
Manuscripts

Patterns of genomic diversification reflect differences in life history and reproductive biology
between figs (*Ficus*) and the stone oaks (*Lithocarpus*)

submitted by

Chai-Shian Kua^{1,2} and Charles H. Cannon^{1,2}

¹Key Lab in Tropical Ecology, Xishuangbanna Tropical Botanical Garden, Menglun, Yunnan
666303 China.

² Current address: The Center for Tree Science, The Morton Arboretum, 4100 Illinois Route 53,
Lisle, IL 60532

Corresponding authors: ccannon@mortonarb.org, ckua@mortonarb.org

Abstract

One of the remarkable aspects of the tremendous biodiversity found in tropical forests is the wide range of evolutionary strategies that have produced this diversity, indicating many paths to diversification. We compare two diverse groups of trees with profoundly different biologies to discover whether these differences are reflected in their genomes. *Ficus* (Moraceae), with its complex co-evolutionary relationship with obligate pollinating wasps, produces copious tiny seeds which are widely dispersed. *Lithocarpus* (Fagaceae), with generalized insect pollination, produce large seeds that are poorly dispersed. We hypothesize that these different reproductive biologies and life history strategies should have a profound impact on the basic properties of genomic divergence within each genus. Using shallow whole genome sequencing for 6 *Ficus* species, 7 *Lithocarpus* species, and 3 outgroups, we examined overall genomic diversity, how it is shared among the species within each genus, and the fraction of this shared diversity which agrees with the major phylogenetic pattern. Substantially larger fraction of the genome is shared among *Lithocarpus* species, a considerable amount of this shared diversity was incongruent with the general background history of the genomes, and each fig species possessed a substantially larger fraction of unique diversity than *Lithocarpus*.

Keywords

Tropical Biodiversity, Reference-free, comparative genomics, *kmers*, *Ficus*, Fagaceae, Moraceae, *Lithocarpus*, *Castanopsis*, *Trigonobalanus*.

Introduction

Trees in tropical forests are very diverse (Slik *et al.* 2015) and many plant groups with different life histories and reproductive biologies have diversified in the equatorial tropics, indicating that many evolutionary strategies promote diversification. Groups with both generalized and specialized pollination systems for example have diversified (Van Steenis 1950). Most tree species can be found in the same forest with at least two closely-related species and the most diverse groups can contribute up to 45 species within a single watershed or sample area in a forest (Cannon and Lerda 2015). A major consequence of high levels of species diversity within a small geographic area is the tremendous opportunity for interspecific gene flow among congeneric taxa and the probability of these taxa evolving as part of a suite of partially interfertile species that interact over long periods of time and a large geographic area, also known as a syngameon (Grant 1971). Differences in life history strategies and reproductive biologies should have an impact on whether species in a particular group of trees interact as a syngameon and how important a role interspecific gene flow plays in the genomic diversification of the group. For example, highly specialized pollination mechanisms within a tree group of sympatric species should act as a stronger reproductive isolation mechanism, thus reducing interspecific gene flow and leading to faster genomic divergence with a clearer phylogenetic signal, when compared to a tree group with a generalized pollination system.

Here, we examine two genera of tropical tree with substantially different life histories and reproductive biologies: the figs (*Ficus*) and the stone oaks (*Lithocarpus*). Both groups are diverse with several hundred species but they have obviously diversified following very evolutionary pathways (Fig 1). In this study, we compare whole genome shallow sequence

(wgss) for 15 species of tropical tree (Table S1). Sampling focused on these two genera but included outgroups. This “wgss” approach is feasible because genome sizes in most angiosperm trees are relatively small (Ohri *et al.* 2004, Petit and Hampe 2006, Chen *et al.* 2014), generally less than a million megabases. Using a reference-free approach based entirely upon the distribution of short *kmers* across the sample genomes (Kua *et al.* 2012), we compare basic whole genomic properties of these two groups to examine how these differing life histories impact genomic diversification. We predict that genomic diversification in stone oaks, given the combination of their life history and reproductive biology, will result in a substantially greater fraction of shared diversity among species and a larger fraction of this diversity will be incongruent with the general background phylogenetic patterns than in the figs.

Materials and methods

In this analysis, we compare two genera of tropical trees that have achieved significant levels of species diversity through substantially different life history strategies and reproductive biologies (Fig. 1). The few exemplar species included in this analysis are representative of the diversity and evolutionary history found within these genera in the Asian tropics. The figs have reached greater levels of overall species diversity with estimates of over 800 species worldwide while the stone oaks are confined to the East Asian tropics where over 200 species are found. In any one location, figs are typically more diverse (Cannon and Ler dau 2015) and fill more ecological niches than the stone oaks (C. Cannon, pers. obs.). Both genera have similar estimates for their age based upon fossil record and divergence times given molecular methods. The genus *Ficus* is roughly 65 million years old (confidence interval: 33-94 MYBP) while the genus *Lithocarpus* is roughly 51 million years old (confidence interval: 43-60 MYBP) (Hedges

et al. 2006). Two of the *Ficus* species are known to be closely related (*Ficus altissima* and *F. microcarpa*) and represent a more recent divergence event than any present among the stone oak species sampled here. The fig genome is slightly less than half the size of the stone oak genome, as referenced from the Kew C-value database (<http://data.kew.org/cvalues/>).

Ficus (Moraceae)

We included 8 whole genomic shallow sequence datasets in the fig analysis: 6 species of *Ficus* (Moraceae : FA- *Ficus altissima*, FM - *Ficus microcarpa*, FL- *Ficus langkokensis*, FT- *Ficus tinctoria*, FR- *Ficus racemosa*, FV- *Ficus vasculosa*) and 2 outgroups (Fagaceae: CI- *Castanopsis indica*; Fabaceae: IB- *Intsia bijuga*). A portion of the *ficus* data for this study have been archived at the NCBI Short Read Archive under the accession number SRP001298. Members from the genus *Ficus* are characterized by an unusual reproductive structure ('synconium') and a complex co-evolutionary relationship with highly-specialized pollinating wasps (Weiblen 2002, Ronsted *et al.* 2008). The relationship between plant and pollinator is completely obligate, largely symbiotic, fairly specific, and supports a multi-trophic microcosm of ecological and evolutionary of hyper-parasites, beneficial and parasitic fungi, and cheaters. Figs have also been long recognized as keystone fruit resources for a wide range of vertebrates in tropical forests, as the plant populations must flower asynchronously to maintain the obligate pollinator populations, thus maintaining a steady supply of fruit, albeit of generally low quality. Their seeds are minute and easily dispersed by a wide range of organisms, from vertebrates to ants. This genus of plant have been very successful throughout the tropics, being distributed globally and having diversified into hundreds of species. The fig 'fruit' is actually a very specialized type of inflorescence, consisting of a flat and wide floral receptacle enclosing tens to

hundreds of tiny, sessile flowers, termed a 'synconium'. This receptacle itself forms the fruit-like structure, with an orifice on the top. Inside lay the flowers, and, after fertilization, the fruits and seeds. Figs can only be pollinated by female Agaonid wasps that oviposit inside the fig cavity, and this mutualism is a model system for studies of co-evolution (Weiblen 2002).

Lithocarpus (Fagaceae)

We included 9 whole genomic shallow sequence datasets in the stone oak analysis: 7 species of *Lithocarpus* (Fagaceae : LB - *Lithocarpus balansea*, LC - *Lithocarpus calolepis*, LF - *Lithocarpus fenestratus*, LG - *Lithocarpus grandifolius*, LH - *Lithocarpus hancei*, LR - *Lithocarpus craibianus*, LX - *Lithocarpus xylocarpus*) and 2 outgroups (Fagaceae: TD - *Trigonobalanus doichangensis*; Moraceae:FA- *Ficus altissima*). Some of the Fagaceae data for this study have been archived at the NCBI Short Read Archive under the accession number SRP001298.

The Fagaceae family, including the temperate oaks, beeches, and chestnuts, is probably one of the better known and studied group of tree, from a genetic standpoint (Plomion *et al.* 2016). Common throughout the middle latitudes of the northern hemisphere, the family also has several genera confined to the Asian tropics, from eastern India extending into the Southeast Asian archipelago to the island of Papua (Soepadmo 1972). These groups, including the stone oaks (*Lithocarpus*), tropical chestnuts (*Castanopsis*), and the Doichang trig-oak (*Trigonobalanus*), are generally old-growth forest specialists and have a generalized insect pollination system, unlike the rest of the family which is wind-pollinated. Little difference exists among species in floral morphology and the pollen can only be distinguished at the genus level.

Fagaceae has infrequent synchronous fruiting, which produces large hard-shelled nuts

(Cannon and Manos 2000). The often large woody acorns produced by the stone oaks contain a single large seed in each fruit are only be dispersed by several larger vertebrates. They are largely incapable of dispersing over large bodies of water and Wallace's Line has a large impact on their distribution. Strong geographic patterns exist in the genetic variation, frequently overwhelming taxonomic patterns so that sympatric heterospecifics are more likely to share haplotypes than allopatric conspecifics (Cannon and Manos 2003). The stone oaks have also diversified into hundreds of species, while the Doichang trig-oak is an ancient relictual species found in scattered and highly endemic populations. This species has persisted for >50 million years without diversifying (Forman 1964).

DNA extraction and sequencing approach

Fresh leaf material was used for DNA extraction. Two types of commercially available DNA extraction kits were used for most leaf samples: Qiagen Plant Dneasy Extraction Kits (Qiagen, Cat#69104) and MN NucleoSpin Plant DNA Extraction Kit (MN, cat#740770.50).

The DNA samples were visualized and quantified on a check gel before shipment to the sequencing facility at the Michael Smith Cancer Institute in British Columbia, Canada. At Canada's Michael Smith Genome Sciences Centre sequencing facility, the genomic DNA samples were sonicated for 10min and run on a 12% PAGE. The 400bp DNA fraction was excised and eluted from the gel slice overnight at 4°C in 300 µL of elution buffer [5:1, LoTE buffer (3 mM Tris-HCl, pH 7.5, 0.2 mM EDTA)-7.5 M ammonium acetate] and purified using a QIAquick purification kit (Qiagen, Cat#28104). Paired-end (PET) sequencing libraries were constructed using Illumina genomic DNA prep kit by following company protocols (Illumina, cat# FC-102-1002). Clusters were generated on the Illumina cluster station (Illumina, cat# FC-

103-1002) and sequence was run on Illumina 1G Analyzer following the manufacturer's instructions (Illumina, cat# FC-104-1003).

Ref-Free Analysis

The two separate data sets for *Ficus* and *Lithocarpus* were each analyzed using the Reference free pipeline (Kua *et al.* 2012) with the following parameters: *kmer* sizes of 17, 21 and 25 base pair, a threshold frequency of 3 for each *kmer* to be included in the analysis, and jackknife sampling of 10%. We performed ten jack-knifing replicates for each data set and present the mean value for these replicates. The analysis presented here is based upon the shared *kmers* table generated during the first phase of the Ref-free analysis (Fig. 2). The *kmers* were classified according to the taxa which shared them and whether the group of shared taxa was congruent with the phylogenetic tree produced using an assembly-free and alignment-free phylogenetic reconstruction technique (Fan *et al.* 2015). Groups within each analysis were ranked according to the number of *kmers* shared by the members of the group.

Results

No obvious difference is apparent among the two groups in the total fraction of unique *kmer* diversity given the subsampling effort or length of the *kmer* (Fig. 3). During each 10% jack-knife replicate, the total number of *kmers* sampled from each genome, for k=17, 21, and 25 bp respectively, was 98 609 826, 87 149 604, and 75 689 382 for *Ficus* and 116 797 624, 103 054 324, and 89 311 024 for *Lithocarpus* and the fraction of unique *kmers* in relation to the whole genome ranged between one half and one quarter of the subsampled portion of the genome. Therefore, none of the genomes appear to be exceptional or different in their total

genomic diversity of *kmers*, with roughly equivalent levels of *kmer* diversity observed among all species for every *kmer* length.

The fraction of these sampled *kmers* that were then identified as being 'tip' *kmers* (unique to a genome) versus 'shared' *kmers* (shared by at least two genomes) differed considerably between genera (Fig. 3A-B). As expected, the outgroups in each analysis were greatly composed of tip *kmers*, with all of the distantly related taxa from a different family being more than 90% unique and the trig-oak was still more than 75% unique, reflecting its deep separation from the rest of the family (Manos *et al.* 2008). Within each genus, the patterns differed considerably between the two analyses, with each *Ficus* species being composed of a substantially greater fraction of tip *kmers* than shared *kmers* than the Fagaceae species. The more basal *Ficus* species possess >70% unique or private *kmer* diversity while the two closely related species, *Ficus altissima* and *Ficus microcarpa*, sharing more than 50% of their *kmers* with other genomes. In the stone oak analysis, all of the genomes shared roughly 70% of their *kmers* and possessed a much smaller fraction of unique or private *kmer* diversity. Additionally, the fraction of the genome which is incongruent with the major phylogenetic pattern in the genomes is substantially greater in among the stone oaks species (~50%) than among fig species (~25%).

Discussion

The substantial differences in life history and reproductive biology between figs and stone oaks (Fig. 1) appear to have a profound effect on general patterns of genomic diversification among the species (Fig. 4A-B). The tight relationship between fig plant and its obligate pollinating wasp (Janzen 1979, Weiblen 2002), the prolific production of seeds and their

high dispersibility should allow these species to diverge more rapidly than the generalized insect pollination and poorly dispersed seeds of the stone oaks. This hypothesis is strongly supported by the fact that a very large fraction ($>70\%$) of each fig genome was unique. Even between two species (*Ficus altissima* and *F. microcarpa*) that are known to be quite closely related, a substantial amount of genomic divergence is apparent. Conversely, the stone oaks share a much greater fraction of their genomes with even distantly related species. Additionally, as would be expected, given the greater opportunity for interspecific gene flow among the stone oaks (Cannon and Manos 2003, Manos *et al.* 2008), the fraction of the genomic *kmer* diversity which is incongruent with the major phylogenetic pattern is much greater than in the figs.

These general characteristics of genomic diversity and divergence among these two major tropical tree groups strongly support the assumption that speciation and phenotypic diversification among tropical trees can occur through profoundly different modes and mechanisms. While the figs are more species rich than the stone oaks and have a global geographic distribution and their evolutionary strategies allow them to diverge more rapidly and with less interspecific gene flow, the stone oaks are also a successful tree group in the Asian tropics and have produced a substantial number of species, occupy a range of habitats, indicating that their strategies are also successful. Both specialized and generalized evolutionary approaches can result in high levels of phenotypic diversity and ecological success through apparently profoundly different modes and patterns of genomic diversification.

The reference-free approach (Kua *et al.* 2012) to the direct comparison of whole genome shallow sequencing is a powerful way to study organisms for which genomic resources are limited. Firstly, the sequencing of the whole genome avoids the possible biases introduced through various reduced representation techniques, like RAD-seq (Hoban *et al.* 2016). Secondly,

because the bulk of the analysis is performed prior to any attempt at assembly, the necessary coverage has been shown to be only 5-8x (Fan *et al.* 2015). In this study, we only utilize the initial part of the overall analysis and examine the broad characteristics of genomic diversity and how it relates to incongruence with the overall phylogenetic history of the genome. We have shown in our previous publication that the partitioning of the *kmers* into their relevant shared groups allows a type of reduced representation that is focused on the genomes pertinent to the research question. These target *kmers* can then be used to extract the associated reads and the localized assembly of only those small portions of the genome identified as being associated with the overall research question. The localized assemblies produce single nucleotide polymorphisms and longer ‘hot-spots’ of mutation that were verified using 174 chloroplast genomes. An additional aspect of this analysis relevant to the large degree of incongruence in the stone oak genomes would be the identification of *kmers* and genomic regions associated with the ‘minor’ phylogenetic histories which contradict the general background history of the genome. These minor histories could represent the more interesting evolutionary events in these groups, when dominant patterns of gene flow and species isolation barriers were altered probably due to major changes in macro-evolutionary processes affecting the groups.

Conclusion

This fundamental comparison of genomic diversity and divergence between two major tropical tree groups, which have each produced a large number of species through different life history and reproductive strategies, reveals that genomic divergence among the fig species is considerably greater than among the stone oak species. Additionally, the stone oak genomes

exhibit a substantially larger degree of conflicting evidence about the phylogenetic history of the species than observed in the figs. These two pieces of evidence agree with the hypothesized impact that their biologies should have on overall genomic divergence. These results can be used to pinpoint the elements that represent important ‘minor’ phylogenetic relationships, where reticulate evolution played a significant role in the evolution of each of these groups.

One of the main challenges facing the application of genomics to the study of tropical biodiversity is the basic lack of relevant completed reference genomes and knowledge about fundamental genomic diversity and how genomes differ. Our reference-free approach allows the direct comparison of whole genome datasets prior to any assembly and the characterisation of genomic divergence and identity. The method also highlights those portions of the genome which are associated with the basic questions and hypotheses we formulate about a particular group and to create more meaningful questions. Given the low cost of the next gen sequencing (on a per base basis), a ‘model organism’ approach is no longer limited to a single species but can instead examine numerous closely related species, as in a “model group”, tackling one of the central elements of tropical biodiversity. Many of the questions facing tropical biologists will require the kind of deep insight and understanding provided by genomic scale studies.

Authors' contributions

Conceived and designed the experiments: CHC CSK. Performed the experiments and analysis: CSK. Wrote the paper: CHC CSK.

Figure captions

Figure 1. Life history and reproductive traits of figs and stone oaks and their predicted impact on genomic diversification within each group.

Figure 2. Flowchart of comparative analysis of genomic data. First, all *kmers* in the original database of next-gen DNA sequence are discovered for each genome (*k* typically ranges from 17 to 25 base pairs, the optimal length varying according to relatedness of species). Second, the individual lists of *kmers* are merged into a table of presence/absence data across all genomes and the group of species sharing the *kmers* are identified. Finally, this table is simplified into a sorted list of groups of species and the number of *kmers* shared by that group (the numbers included in the table are for illustration only). Full analytical approach is described in Kua *et al.* 2012.

Figure 3. Mean fraction of unique *kmers* sampled from each genome during each 10% jackknife analysis for different *kmer* lengths. *Lithocarpus* = green. *Ficus* = red. Each point represents the mean for each species while the line indicates the average for all species in each genus.

Figure 4. Genomic fractions in each species of *kmers* according to whether they were unique to a genome (white) ; shared with at least one other species and congruent with the major phylogenetic pattern (green) ; or shared with at least one other species and incongruent with the major phylogenetic pattern (red). A) fig analysis. B) stone oak analysis. Phylogeny for each analysis, including branch lengths, was adapted from Fan *et al.* 2015. Outgroups are shown at

the bottom of each panel.

Draft

References

- Cannon, C.H., and Lerda, M. 2015. Variable mating behaviors and the maintenance of tropical biodiversity. *Front. Genet.* **6**: 183.
- Cannon, C.H., and Manos, P.S. 2000. The Bornean *Lithocarpus* Bl. section *Synaedrys* (Lindl.) Barnett (Fagaceae): its circumscription and description of a new species. *Bot. J. Linn. Soc.* **133**(3): 343–357.
- Cannon, C.H., and Manos, P.S. 2003. Phylogeography of the Southeast Asian stone oaks (*Lithocarpus*). *J. Biogeogr.* **30**: 211–226.
- Chen, S.-C., Cannon, C.H., Kua, C.-S., Liu, J.-J., and Galbraith, D.W. 2014. Genome size variation in the Fagaceae and its implications for trees. *Tree Genet. Genomes*. doi:10.1007/s11295-014-0736-y.
- Fan, H., Ives, A.R., Surget-Groba, Y., and Cannon, C.H. 2015. An assembly and alignment-free method of phylogeny reconstruction from next-generation sequencing data. *BMC Genomics* **16**: 522.
- Forman, L.L. 1964. *Trigonobalanus*, a New Genus of Fagaceae, with Notes on the Classification of the Family. *Kew Bull.* **17**(3): 381–396. [Springer, Royal Botanic Gardens, Kew].
- Grant, V. 1971. *Plant Speciation*. Columbia University Press.
- Hedges, S.B., Dudley, J., and Kumar, S. 2006. TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* **22**(23): 2971–2972.
- Hoban, S., Kelley, J.L., Lotterhos, K.E., Antolin, M.F., Bradburd, G., Lowry, D.B., Poss, M.L., Reed, L.K., Storfer, A., and Whitlock, M.C. 2016. Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. *Am. Nat.* **188**(4): 379–397.
- Janzen, D.H. 1979. How To Be A Fig. *Annu. Rev. Ecol. Syst.* **10**: 13–51.
- Kua, C.-S., Ruan, J., Harting, J., Ye, C.-X., Helmus, M.R., Yu, J., and Cannon, C.H. 2012. Reference-Free Comparative Genomics of 174 Chloroplasts. *PLoS One* **7**(11): e48995. [accessed 22 November 2013].
- Manos, P.S., Cannon, C.H., and Oh, S.H. 2008. Phylogenetic relationships and taxonomic status of the paleoendemic Fagaceae of western North America: recognition of a new genus *Notholithocarpus*.

- Madrono **55**(3): 181–190.
- Ohri, D., Bhargava, A., and Chatterjee, A. 2004. Nuclear DNA amounts in 112 species of tropical hardwoods - new estimates. *Plant Biol.* **6**: 555–561.
- Petit, R.J., and Hampe, A. 2006. Some evolutionary consequences of being a tree. *Annu. Rev. Ecol. Evol. Syst.* **37**: 187–214.
- Plomion, C., Aury, J.-M., Amselem, J., Alaeitabar, T., Barbe, V., Belser, C., Bergès, H., Bodénès, C., Boudet, N., Boury, C., Canaguier, A., Couloux, A., Da Silva, C., Duplessis, S., Ehrenmann, F., Estrada-Mairey, B., Fouteau, S., Francillon, N., Gaspin, C., Guichard, C., Klopp, C., Labadie, K., Lalanne, C., Le Clainche, I., Leplé, J.-C., Le Provost, G., Leroy, T., Lesur, I., Martin, F., Mercier, J., Michotey, C., Murat, F., Salin, F., Steinbach, D., Faivre-Rampant, P., Wincker, P., Salse, J., Quesneville, H., and Kremer, A. 2016. Decoding the oak genome: public release of sequence data, assembly, annotation and publication strategies. *Mol. Ecol. Resour.* **16**(1): 254–265.
- Ronsted, N., Weiblen, G.D., Savolainen, V., and Cook, J.M. 2008. Phylogeny, biogeography, and ecology of *Ficus* section *Malvanthera* (Moraceae). *Mol. Phylogenet. Evol.* **48**(1): 12–22.
- Slik, J.W.F., Arroyo-Rodríguez, V., Aiba, S.-I., Alvarez-Loayza, P., Alves, L.F., Ashton, P., Balvanera, P., Bastian, M.L., Bellingham, P.J., van den Berg, E., Bernacci, L., da Conceição Bispo, P., Blanc, L., Böhning-Gaese, K., Boeckx, P., Bongers, F., Boyle, B., Bradford, M., Brearley, F.Q., Breuer-Ndoundou Hockemba, M., Bunyavejchewin, S., Calderado Leal Matos, D., Castillo-Santiago, M., Catharino, E.L.M., Chai, S.-L., Chen, Y., Colwell, R.K., Chazdon, R.L., Robin, C.L., Clark, C., Clark, D.B., Clark, D.A., Culmsee, H., Damas, K., Dattaraja, H.S., Dauby, G., Davidar, P., DeWalt, S.J., Doucet, J.-L., Duque, A., Durigan, G., Eichhorn, K.A.O., Eisenlohr, P.V., Eler, E., Ewango, C., Farwig, N., Feeley, K.J., Ferreira, L., Field, R., de Oliveira Filho, A.T., Fletcher, C., Forshed, O., Franco, G., Fredriksson, G., Gillespie, T., Gillet, J.-F., Amarnath, G., Griffith, D.M., Grogan, J., Gunatilleke, N., Harris, D., Harrison, R., Hector, A., Homeier, J., Imai, N., Itoh, A., Jansen, P.A., Joly, C.A., de Jong, B.H.J., Kartawinata, K., Kearsley, E., Kelly, D.L., Kenfack, D., Kessler, M.,

- Kitayama, K., Kooyman, R., Larney, E., Laumonier, Y., Laurance, S., Laurance, W.F., Lawes, M.J., Amaral, I.L. do, Letcher, S.G., Lindsell, J., Lu, X., Mansor, A., Marjokorpi, A., Martin, E.H., Meilby, H., Melo, F.P.L., Metcalfe, D.J., Medjibe, V.P., Metzger, J.P., Millet, J., Mohandass, D., Montero, J.C., de Morisson Valeriano, M., Mugerwa, B., Nagamasu, H., Nilus, R., Ochoa-Gaona, S., Onrizal, Page, N., Parolin, P., Parren, M., Parthasarathy, N., Paudel, E., Permana, A., Piedade, M.T.F., Pitman, N.C.A., Poorter, L., Poulsen, A.D., Poulsen, J., Powers, J., Prasad, R.C., Puyravaud, J.-P., Razafimahaimodison, J.-C., Reitsma, J., Dos Santos, J.R., Roberto Spironello, W., Romero-Saltos, H., Rovero, F., Rozak, A.H., Ruokolainen, K., Rutishauser, E., Saiter, F., Saner, P., Santos, B.A., Santos, F., Sarker, S.K., Satdichanh, M., Schmitt, C.B., Schöngart, J., Schulze, M., Suganuma, M.S., Sheil, D., da Silva Pinheiro, E., Sist, P., Stevart, T., Sukumar, R., Sun, I.-F., Sunderland, T., Sunderand, T., Suresh, H.S., Suzuki, E., Tabarelli, M., Tang, J., Targhetta, N., Theilade, I., Thomas, D.W., Tchouto, P., Hurtado, J., Valencia, R., van Valkenburg, J.L.C.H., Van Do, T., Vasquez, R., Verbeeck, H., Adekunle, V., Vieira, S.A., Webb, C.O., Whitfeld, T., Wich, S.A., Williams, J., Wittmann, F., Wöll, H., Yang, X., Adou Yao, C.Y., Yap, S.L., Yoneda, T., Zahawi, R.A., Zakaria, R., Zang, R., de Assis, R.L., Garcia Luize, B., and Venticinque, E.M. 2015. An estimate of the number of tropical tree species. *Proc. Natl. Acad. Sci. U. S. A.* **112**(24): 7472–7477.
- Soepadmo, E. 1972. “Fagaceae” *In* Flora Malesiana: Series I - Spermatophytes. Noordhoff International Publishing.
- Van Steenis, C.G.G.J. (*Editor*). 1950-. Flora Malesiana. Ser. I, Spermatophyta, Flowering Plants. Sijthoff & Noordhoff International Publishers, Alphen aan den Rijn. The Netherlands.
- Weiblen, G.D. 2002. How to be a fig wasp. *Annu. Rev. Entomol.* **47**: 299–330.

Figure 2. Flowchart of comparative analysis of genomic data. First, all kmers in the original database of next-gen DNA sequence are discovered for each genome (k typically ranges from 17 to 25 base pairs, the optimal length varying according to relatedness of species). Second, the individual lists of kmers are merged into a table of presence/absence data across all genomes and the group of species sharing the kmers are identified. Finally, this table is simplified into a sorted list of groups of species and the number of kmers shared by that group (the numbers included in the table are for illustration only). Full analytical approach is described in Kua et al. 2012.

	<i>Ficus</i> (Figs)	<i>Lithocarpus</i> (stone oaks)
Mode of sexual reproduction	Frequent asynchronous fruiting	Infrequent synchronous fruiting
Pollination biology	Symbiotic relationship with specific wasp pollinator	Generalized diverse insect pollination
Spatial scale of gene flow	Copious tiny seeds widely dispersed by many animals	Few large seeds poorly dispersed by large vertebrates
Predicted impact on genomic diversification	Rapid divergence Strong phylogenomic signal	Slow divergence Weak phylogenomic signal

Figure 1. Life history and reproductive traits of figs and stone oaks and their predicted impact on genomic diversification within each group.

A

B

Table S1. The main characteristics of the genomic data in the Fagaceae (1A) and Ficus (1B) reference fr. Each taxon is shown on a row. "Total dataset" indicates the number of "reads" available and the total diversity in the original dataset. "Sample d kmers" indicates the number of " kmers" and its percentage of the original dataset, making each dataset equivalent in sampling effort and using a 10% jack-knife.

Taxa	Total dataset		Sampled	
	Reads	total K	kmers	% total
<i>Litnocarpus dalansea</i> (LB)	1030543240	34358250	3172964	9.2%
<i>Litnocarpus caliolepis</i> (LC)	1240930988	40954034	3806034	9.3%
<i>Litnocarpus jenestratus</i> (LF)	6910791008	127758294	2987630	2.3%
<i>Litnocarpus granatjoius</i> (LG)	1177138160	38999224	3253028	8.3%
<i>Litnocarpus nancei</i> (LH)	1554847466	57728450	2643406	4.6%
<i>Litnocarpus craibianus</i> (LK)	6198226066	116996242	3046807	2.6%
<i>Litnocarpus xylocarpus</i> (LX)	1602524048	52726770	3178643	6.0%
<i>Irigonobalanus doichangensis</i> (LD)	1775331926	65219558	2543894	3.9%
<i>Ficus altissima</i> (FA)	1320075126	43258938	2507767	5.8%
<i>Ficus langkokensis</i> (FL)	1204200066	39446760	3211127	8.1%
<i>Ficus microcarpa</i> (FMI)	871496044	28650556	2363138	8.2%
<i>Ficus racemosa</i> (FR)	6421422160	118363636	2820309	2.4%
<i>Ficus tinctoria</i> (FI)	2797562116	92475558	2524353	2.7%
<i>Ficus vasculosa</i> (FV)	6836667968	125796860	2916391	2.3%
<i>Intsia bijuga</i> (IB)	6384780782	120660440	2451372	2.0%

tree analysis
kmer (21 bp)
e sampled out

Draft