

**HYPOTHESES**

# On the hunt for the alternate host of *Hemileia vastatrix*

Athina Koutouleas<sup>1</sup> | Hans Jørgen Lyngs Jørgensen<sup>2</sup> | Birgit Jensen<sup>2</sup> |  
Jens-Peter Barnekow Lillesø<sup>1</sup> | Alexander Junge<sup>3</sup> | Anders Ræbild<sup>1</sup> <sup>1</sup>Department of Geosciences and Natural Resource Management, University of Copenhagen, Frederiksberg C, Denmark<sup>2</sup>Department of Plant and Environmental Sciences and Copenhagen Plant Science Centre, University of Copenhagen, Frederiksberg C, Denmark<sup>3</sup>Faculty of Health and Medical Sciences, Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Copenhagen N, Denmark**Correspondence**Athina Koutouleas, Department of Geosciences and Natural Resource Management, University of Copenhagen, Rolighedsvej 23, DK-1871 Frederiksberg C, Denmark.  
Email: atk@ign.ku.dk**Funding information**

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 727934. The funding body played no role in the design of the study and collection, analysis, and interpretation of data.

**Abstract**

Coffee leaf rust (CLR), caused by the fungal pathogen *Hemileia vastatrix*, has plagued coffee production worldwide for over 150 years. *Hemileia vastatrix* produces urediniospores, teliospores, and the sexual basidiospores. Infection of coffee by basidiospores of *H. vastatrix* has never been reported and thus far, no alternate host, capable of supporting an aecial stage in the disease cycle, has been found. Due to this, some argue that an alternate host of *H. vastatrix* does not exist. Yet, to date, the plant pathology community has been puzzled by the ability of *H. vastatrix* to overcome resistance in coffee cultivars despite the apparent lack of sexual reproduction and an aecial stage. The purpose of this study was to introduce a new method to search for the alternate host(s) of *H. vastatrix*. To do this, we present the novel hypothetical alternate host ranking (HAHR) method and an automated text mining (ATM) procedure, utilizing comprehensive biogeographical botanical data from the designated sites of interests (Ethiopia, Kenya and Sri Lanka) and plant pathology insights. With the HAHR/ATM methods, we produced prioritized lists of potential alternate hosts plant of coffee leaf rust. This is a first attempt to seek out an alternate plant host of a pathogenic fungus in this manner. The HAHR method showed the highest-ranking probable alternate host as *Psychotria mahonii*, *Rubus apetalus*, and *Rhamnus prinoides*. The cross-referenced results by the two methods suggest that plant genera of interest are *Croton*, *Euphorbia*, and *Rubus*. The HAHR and ATM methods may also be applied to other plant-rust interactions that include an unknown alternate host or any other biological system, which rely on data mining of published data.

**KEYWORDS**coffee leaf rust, disease cycle, *Hemileia vastatrix*, hypothetical alternate host ranking

## 1 | INTRODUCTION

The genus *Coffea* is composed of over one hundred species, which grow wild in equatorial Africa and Madagascar (Lashermes, Bertrand, & Ettienne, 2009; McCook, 2006). One of the major diseases threatening coffee production is coffee leaf rust (CLR), caused by the biotrophic rust fungus, *Hemileia vastatrix* Berk. &

Broome (Basidiomycota, Pucciniales) (Berkeley & Broome, 1869; Toniutti et al., 2017). Wild coffee species and *H. vastatrix* have co-evolved for hundreds of years in equatorial Africa, and the fungus was restricted to this continent up until the mid-nineteenth century (McCook, 2006). Today, *H. vastatrix* is able to infect all known cultivated species in the genus *Coffea*, albeit at different levels of severity (McCook, 2006). Since the first significant outbreak in Sri

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

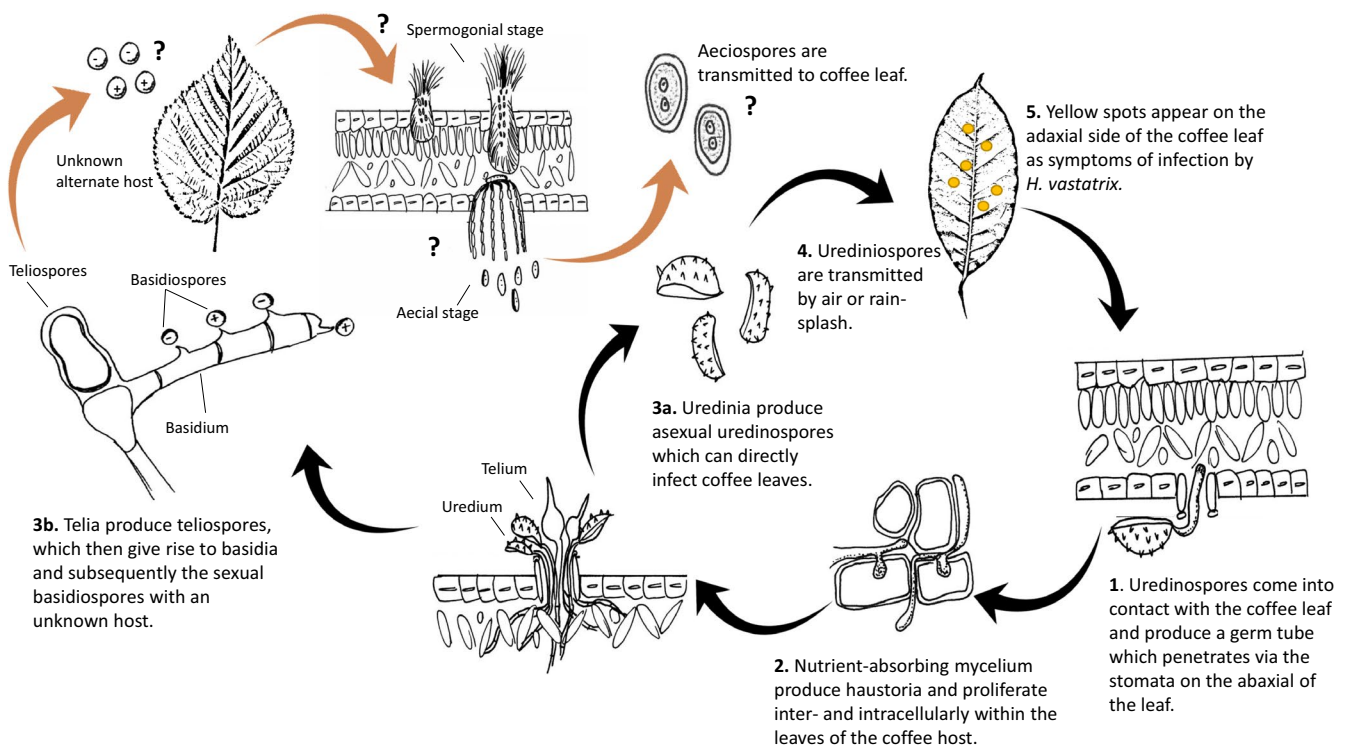
© 2019 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

Lanka (Ceylon) in 1869, almost the entire world's coffee producing zones have reported coffee leaf rust attacks, resulting in up to 40% annual yield losses (Arneson, 2000; Kumar, Sreedharan, Shetty, & Parvatam, 2016; McCook, 2006).

*Hemileia vastatrix* penetrates coffee leaves via the stomatal openings and grows nutrient-absorbing mycelium through the leaf mesophyll. Vibrant bouquet-shaped, orange uredinia and telia are produced on the abaxial side of the coffee leaves (Arneson, 2000; Kumar et al., 2016). Uredinia give rise to urediniospores, which are dikaryotic and the only reported means of propagation for *H. vastatrix* (Arneson, 2000; Carvalho, Fernandes, Carvalho, Barreto, & Evans, 2011) (Figure 1). Dry urediniospores can survive up to 6 weeks on detached plant tissue, but will only germinate again in the presence of rain or heavy dew (Arneson, 2000). Under cool, dry conditions, the telia give rise to the two more elusive spore types: teliospores and subsequently basidiospores (Arneson, 2000; Coutinho, Rijkenberg, & Asch, 1995). Teliospores are two-celled, thick-walled and consist of dikaryotic cells (Schumann & Leonard, 2000). Teliospores produce basidia, which then develop four haploid basidiospores (Arneson, 2000; Coutinho et al., 1995) (Figure 1). In most rust fungi, only the teliospores are capable of long-term survival away from a living host plant (Schumann & Leonard, 2000). By producing both asexual and sexual spore types, rust fungi increase the chance of transmission to multiple hosts (Shattock & Preece, 2000). For this reason, many rusts are observed to have complex disease cycles with different spore types or reproductive structures being defined as either macrocyclic (producing five spore types: spermatia, aeciospores, urediniospores, teliospores, and basidiospores) or microcyclic (species

often lacking aeciospores and urediniospores, with or without spermatia) (Shattock & Preece, 2000). The sexual stage of a rust fungus' life cycle is of particular importance, because it facilitates the rise of new genotypes via recombination (Shattock & Preece, 2000). Despite the long history of CLR and the wide interest of the plant pathology community, critical aspects of the disease cycle of *H. vastatrix* remain unclear (Carvalho et al., 2011). Some have hypothesized that *H. vastatrix* is a heteroecious rust, thus requiring two hosts for the completion of the disease cycle (Gopalkrishnan, 1951; Petersen, 1974). The fact that basidiospores do not re-infect coffee supports this theory (Gopalkrishnan, 1951). Yet, an alternate host of *H. vastatrix* has never been reported. It has been postulated that the basidiospores of *H. vastatrix* are remnants of an earlier rust ancestor and no longer utilized by the fungus (Arneson, 2000; Waller, 1982). However, others argue that the preservation of the basidiospores in the observed disease cycle provides evidence for a viable, alternate host of *H. vastatrix* (Petersen, 1974). Others have speculated that based on Tranzschel's Law (Shattock & Preece, 2000), the alternate host of *H. vastatrix* is an orchid (Rodrigues, 1990).

One of the earliest attempts to re-infect coffee leaves with the "sporidia" (aka. basidiospores) arising from *H. vastatrix* teliospores was described as an "utter failure" (Ward, 1882). Since then, there have apparently been no reports of infection by *H. vastatrix* basidiospores in any plant species. This leads us to ask why this spore type is being produced by the fungus at all? There are examples of autoecious (single host) rust fungi, which can infect the same host with all spore types, such as the macrocyclic rust *Puccinia helianthi*, the causal agent of sunflower rust (Hiratsuka & Sato, 1982). However,



**FIGURE 1** Disease cycle of *Hemileia vastatrix*. The black arrows indicate currently known uredinal stage (modified from Arneson, 2000), and orange arrows indicate the hypothesized aecial stage occurring on an unknown host plant(s)

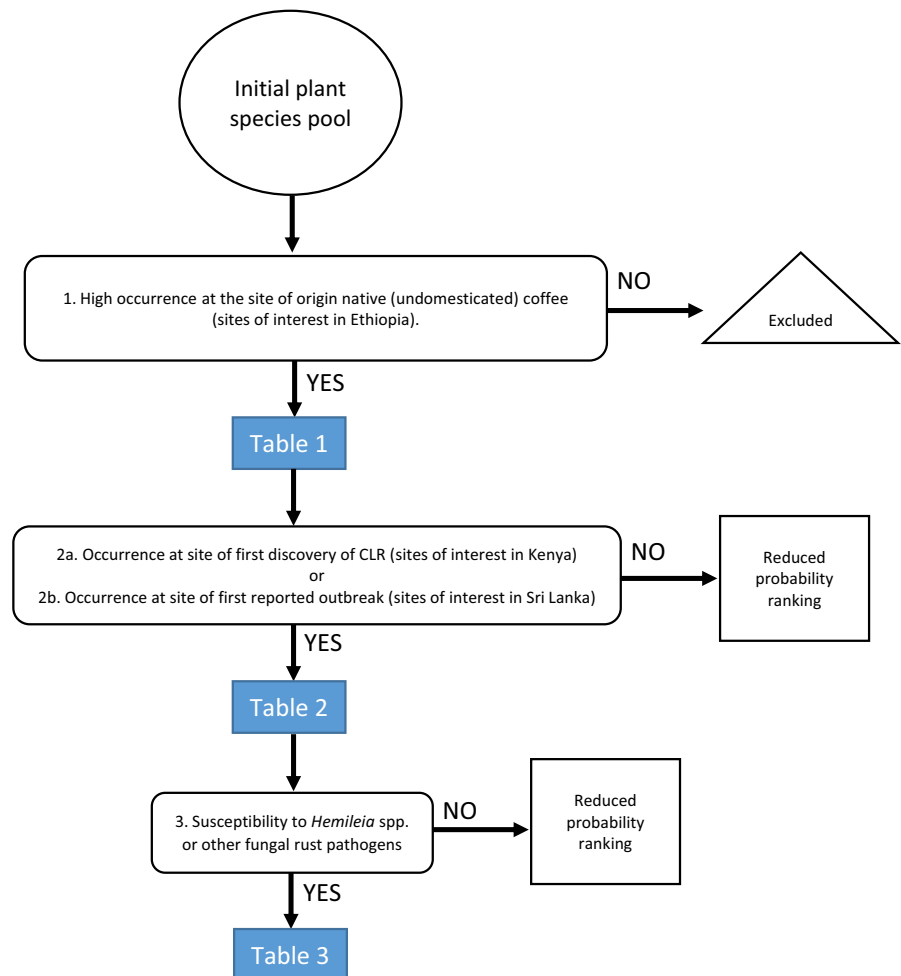
it is most often observed that basidiospores do not infect the same plant species from which they originated (Kolmer, Ordonez, & Groth, 2009; Petersen, 1974). This implies that there is a high likelihood of an unrelated, alternate host, which *H. vastatrix* could infect to produce spermogonia and later aecia to complete the disease cycle (Figure 1).

Furthermore, a historical report by a British expedition to Sri Lanka in 1882 led to specimen collections of “jungle leaves” including palms, dicots, ferns, and grasses that exhibited the characters of *H. vastatrix* being chlorotic yellow, “pin-spots” (Ferguson & Ferguson, 1882). However, upon later scientific examination, no signs of *H. vastatrix* could be confirmed (Ferguson & Ferguson, 1882). To the best of our knowledge, no subsequent studies to search for the possible aecial hosts of *H. vastatrix* have been published. Another possibility is that multiple host species of *H. vastatrix* exist, as with the *Cronartium* species *C. flaccidum* and *C. ribicola*. These rust pathogens have been reported to infect eight diverse host plants from six different families in greenhouse inoculation experiments (Kaitera, Hiltunen, & Hantula, 2017).

Modern coffee breeding and cultivation have led to a continuous evolution exertion on *H. vastatrix* by selection for resistance to CLR in commercial *Coffea* spp. cultivars (Silva, Várzea, Paulo, & Batista, 2018). Today, more than 50 races of *H. vastatrix* are known (Talhinhas et al., 2017). This is an inexplicable evolution for a pathogen that

supposedly only utilizes clonal reproduction (Silva et al., 2018). Some reports have started to emerge, hypothesizing that the different races of *H. vastatrix* are the result of cryptosexuality, that is, the occurrence of hidden sexual reproduction within the urediniospores (Carvalho et al., 2011). However, these new findings would not explain the ability of *H. vastatrix* to produce basidiospores from the teliospores. Another hypothesis relating to the CLR outbreaks in Central America are based on primary host density (Burdon & Chilvers, 1982). This implies that the epidemics of CLR occurred due to the thousands of coffee trees planted in succession within coffee growing regions in the central Americas. This would exclude the need for an alternate host in order for *H. vastatrix* to proliferate and spread, as the primary host is densely planted and highly accessible to the pathogen. However, this hypothesis does not allow for new variation of the pathogen, but merely maintenance of the clonal propagation of *H. vastatrix*.

The plant pathology community have adopted a somewhat ad hoc approach to identify alternate host plant species, whereby such species are often found serendipitously in disease-prone environments using a not always structured approach (McDonald, Richardson, Zambino, Klopfenstein, & Kim, 2006; Rodriguez-Algaba, Walter, Sørensen, Hovmøller, & Justesen, 2014). This has been the case for the alternate host of *Puccinia striiformis* (causal organism of yellow (or stripe) rust of cereal crops), which was recently identified as *Berberis chinensis* after



**FIGURE 2** Decision tree used to elude the hypothetical alternate host ranking (HAHR) of *Hemileia vastatrix*

having been unknown for a century (Jin, Szabo, & Carson, 2010). Later, *Berberis vulgaris* was also identified as an alternate host of *P. striiformis* (Rodriguez-Algaba et al., 2014). Since *B. vulgaris* has been known as the aecial host of *Puccinia graminis* for decades, this may have masked the discovery of *B. vulgaris* as the alternate host of *P. striiformis*. These case studies show the difficulty in determining alternate host species in the context of plant pathology.

Here we present the hypothetical alternate host ranking (HAHR) and automated text mining (ATM) methods to address this gap in knowledge based on a series of assumptions relating to the disease biology of this given pathogen. Our use and integration of comprehensive geographical flora data mapping is novel to traditional plant pathology publications. We believe that this new approach will encourage more multidisciplinary collaborations and hypothesis generation for future studies in this area among plant pathologists and botanists.

## 2 | METHOD

We formulated the so-called HAHR method in order to create ranked lists of plant species, which could be likely alternate host(s) of *H. vastatrix*. This HAHR takes the form of a decision tree (Figure 2). To begin with, an initial plant species pool was compiled, consisting of 377 different species or genera collated from different sources relating to flora mapping at one of the approximate sites of origin of native (undomesticated) coffee (south-western highlands of Ethiopia). The sources used included primary literature, online flora databases, and potential vegetation maps. Relevant primary literature was retrieved through a database search using the Web of Science (<https://webofknowledge.com/>) and Google Scholar (<https://scholar.google.dk/>) as of 11 June 2019. The following search terms were used both singularly and in combination: "coffee," "coffea," "flora mapping," "vegetation," "origin," "Ethiopia," "Kenya," and "Sri Lanka." Plants species and genera listed in over 40 primary sources were then collected and arranged in an MS Excel spreadsheet. The filter function was used to rank plant species or genera according to their co-occurrence at the site of first discovery of *H. vastatrix* (Lake Kenya region) (Ferreira & Boley, 1991; Waller, 1982) and/or the site of first reported outbreak of CLR (Sri Lanka [Ceylon]) (Berkeley & Broome, 1869). Previous reports of susceptibility to *Hemileia* spp. or other rust pathogens were prioritized in the final ranking. Plant names and authors were verified by The Taxonomic Name Resolution Service (iPlant Collaborative) (available from: <http://tnrs.iplantcollaborative.org>). Where the name or author match score was less than 100%, The Plant List ([www.theplantlist.org](http://www.theplantlist.org)) was used to crosscheck species. A high-, medium-, and low-ranking list of plant species was then produced based on this HAHR. The rationale for each ranking criterion is described below.

### 2.1 | Co-occurrence with native (undomesticated) *Coffea* spp. at the site of origin

We started by determining the initial plant species pool based on flora mapping studies performed in co-occurrence of wild *Coffea*

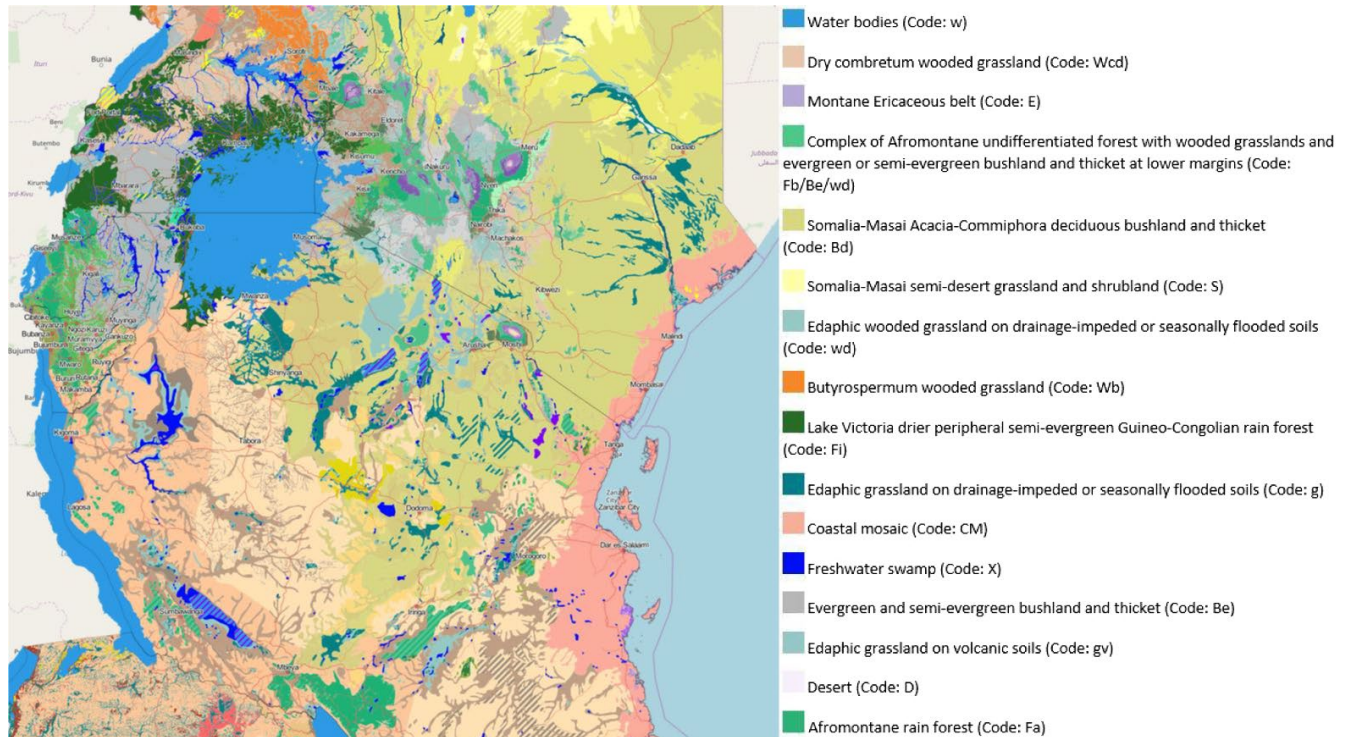
spp., specifically in the south-western highlands of Ethiopia (Gole, 2003; Kelbessa & Soromessa, 2008; Nune, 2008; Schmitt, 2006; Senbeta & Denich, 2006; Tadesse & Nigatu, 1996). Most of the literature found was based on either *Coffea arabica* or nondefined species of wild coffee. As it is by no means certain which species of coffee that *H. vastatrix* co-evolved with, there is an inherent assumption of origin with *C. arabica* or other unknown wild relatives in our method, based on the literature that was available.

Potential natural vegetation (PNV) maps of Ethiopia and Kenya were also used to compile the initial plant species pool (Table S1) (van Breugel et al., 2015). PNV maps are defined to illustrate vegetation that would persist under the current climatic conditions without human intervention (van Breugel et al., 2015). The Keffa and Sidamo regions (where the Geba-Dogi, Berhane-Kontir, Boginda-Yeba, and Harena forest areas lie) have been repeatedly recognized as one of the most probable origins of wild *Coffea* species (Gole, 2003; Meyer, 1965; Schmitt, 2006; Senbeta & Denich, 2006). These regions were collectively assumed as a *site of origin* (or the sites of interest in Ethiopia). According to PNV maps, the sites of interest in Ethiopia were postulated as consisting of either "Complex of Afromontane undifferentiated forest" with "wooded grasslands" or "evergreen or semi-evergreen bushland and thicket at lower margins" (Figure 3, code: Fb/Be/wd, Kaffa region, Ethiopia) and/or "Afromontane rain forest" (Figure 3, code: Fa, Sidamo region, Ethiopia). The Global Biodiversity Information Facility (<https://www.gbif.org/>) was used to cross-reference the primary literature with the PNV map. Plants species, which fulfilled these criteria, are listed in Table 1.

### 2.2 | Co-occurrence at the site of first discovery of CLR

Subsequently, we analyzed flora mapping at the site of first discovery of CLR. The earliest discovery of CLR was in the Lake Victoria region of Kenya in 1861 by a British explorer on uncultivated, wild coffee (Ferreira & Boley, 1991; Waller, 1982). Given this, we assumed the Lake Victoria region to be the natural site of first discovery of CLR. Based on the PNV maps, the sites of interest in Kenya were hypothesized to be surrounding either the "Lake Victoria transitional rain forest" (Figure 3, code: Ff), "Lake Victoria drier peripheral semi-evergreen Guineo-Congolian rain forest" (Figure 3, code: Fi), and/or "Afromontane rain forest" (Figure 3, code: Fa) in close proximity to Kisumu, Kenya. The PNV species web map of eastern Africa (<http://vegetationmap4africa.org>) was then used to verify the plant species, which occurred concomitantly with the sites of interest in Ethiopia and Kenya. Species which were classified by the PNV species web map as "characteristic" (those documented on at least half of all the national manifestations of the vegetation type) or "present" (those documented to be characteristic in at least one of the national manifestations of the vegetation type) at the sites of interest in Kenya were considered for further analysis. "Marginal" species (those listed on national documentation, but not categorized as characteristic or present) were excluded from the analysis.





**FIGURE 3** Potential natural vegetation (PNV) map of Ethiopia and Kenya used to define and verify the initial plant species pool (<http://maps.vegetationmap4africa.org/>). Code regions: Ff, Fa, and Fb/Be/wd were classified as the sites of interest in the hypothetical alternate host ranking method

### 2.3 | Co-occurrence at the site of first major outbreak of CLR

The first major outbreak of CLR was reported in Sri Lanka (Ceylon) in 1869 (Berkeley & Broome, 1869). After this, CLR spread around the world in three sequential outbreaks (McCook, 2006). From 1870 to 1920, the disease spread through to the Indian Ocean Basin and the Pacific. From 1950 to 1960, West African coffee production was severely affected by CLR. Finally, from the late 1960s onwards, CLR spread across the coffee producing zones throughout the Americas (McCook, 2006).

Transcontinental air dispersion of *H. vastatrix* urediniospores has been reported, but this spore type is more frequently dispersed locally by rain-splash due to a tendency to adhere strongly to each other, to leaves, and to smooth surfaces (Brown & Hovmøller, 2002; Nutman, Roberts, & Bock, 1960). Thus, the disease severity of CLR has often been associated with heavy rainfall (Waller, 1982). Due to these aspects of spore transmission and the fact that urediniospores have a limited ability to survive on nonliving coffee leaves, we assumed that this first CLR outbreak was the result of the longer-living teliospores being transported to Sri Lanka on dry plant material. We also assumed that this outbreak was enhanced by the presence of a supportive alternate host in Sri Lanka for the basidiospores to infect, in order to facilitate the generation of new virulent races. Primary flora data (Ashton & Gunatilleke, 1987; Ashton et al., 1997) and The Global Biodiversity Information Facility (<https://www.gbif.org/>) were used to cross-reference the plant species or genera present at

the sites of interest in Ethiopia, Kenya, and Sri Lanka. Plants, which fulfilled the criteria outlined in the method in Sections 2.2 and 2.3, are listed in Table 2.

### 2.4 | Susceptibility to *Hemileia* spp. or other rust fungi

Finally, plant species or genera found concomitantly at the sites of interest in Ethiopia, Kenya, and/or Sri Lanka were also investigated for: (a) any known susceptibility to the *Hemileia* rust species reported (Table S2); and (b) susceptibility to other rust pathogens. This was performed by a review of primary literature obtained through a database search using the Web of Science (<https://webofknowledge.com/>) and Google Scholar (<https://scholar.google.dk/>) as of 11 June 2019, plant pathology reference works (Arthur, 1934; Wilson & Henderson, 1966) and the MyCoPortal (<http://mycoportal.org>). We focused on studies or observations concerning the plant species or genera of interest and any associated rust pathogen. The following search terms were used (in combination with the plant species or genus name): “rust,” “fungi,” “*Hemileia*,” or “fungal pathogen.” All fungal names and authors were verified by Species Fungorum (<http://www.speciesfungorum.org/>). Table 3 is made based on the presence of the plant species or genera at the site of origin (Ethiopia) and/or the site of first discovery (Kenya) or first outbreak (Sri Lanka) as well as association with (a) *Hemileia* spp. and (b) other rust fungi. The different rust spore stages were classified according to Hiratsuka and Sato (1982):

**TABLE 1** Low-ranking list of species considered as potential alternate host(s) of *Hemileia vastatrix*, based on co-occurrence with the sites of interest in Ethiopia (shown in alphabetical order)

<i>Acacia abyssinica</i> Hochst. ex Benth.
<i>Acacia brevispica</i> Harms
<i>Acacia gerrardii</i> Benth.
<i>Acacia mellifera</i> (Vahl) Benth.
<i>Acacia senegal</i> (L.) Willd.
<i>Acacia seyal</i> Delile
<i>Acacia sieberiana f. nefasia</i> (Hochst. ex A. Rich.) Roberty
<i>Acacia tortilis</i> (Forssk.) Hayne
<i>Acacia xanthophloea</i> Benth.
<i>Achyranthes aspera</i> L.
<i>Afrocanthium keniense</i> (Bullock) Lantz
<i>Afrocanthium lactescens</i> (Hiern) Lantz
<i>Afrocarpus falcatus</i> (Thunb.) C.N.Page
<i>Agarista salicifolia</i> (Lam.) G.Don
<i>Alangium chinense</i> (Lour.) Harms
<i>Albizia coriaria</i> Oliv.
<i>Albizia grandibracteata</i> Taub.
<i>Albizia gummifera</i> (J.F.Gmel.)C.A.Sm.
<i>Albizia schimperiana</i> Oliv.
<i>Albizia zygia</i> (DC.)J.F.Macbr.
<i>Alchornea hirtella</i> Benth.
<i>Allophylus abyssinicus</i> (Hochst.) Radlk.
<i>Allophylus africanus</i> P.Beauv.
<i>Allophylus ferrugineus</i> Taub.
<i>Anthocleista grandiflora</i> Gilg
<i>Apodytes dimidiata</i> E.Mey. ex Arn.
<i>Balthasaria schliebenii</i> (Melch.) Verdc.
<i>Bersama abyssinica</i> Fresen.
<i>Blighia unijugata</i> Baker
<i>Bridelia brideliifolia</i> (Pax) Fedde
<i>Bridelia micrantha</i> (Hochst.) Baill.
<i>Carapa procera</i> DC.
<i>Casearia battiscombei</i> R.E.Fr.
<i>Cassipourea malosana</i> (Baker) Alston
<i>Cassipourea ruwensorensis</i> (Engl.) Alston
<i>Catha edulis</i> (Vahl) Endl.
<i>Celtis africana</i> Burm.f.
<i>Celtis gomphophylla</i> Baker
<i>Chrysophyllum gorungosanum</i> Engl.
<i>Clausena anisata</i> (Willd.) Hook.f. ex Benth.
<i>Cola greenwayi</i> Brenan
<i>Cordia africana</i> Lam.
<i>Cornus volkensii</i> Harms
<i>Crotalaria agatiflora</i> Schweinf.

(Continues)

**TABLE 1** (Continued)

<i>Croton dichogamus</i> Pax
<i>Croton macrostachyus</i> Hochst. ex Delile
<i>Croton megalocarpus</i> Hutch.
<i>Croton sylvaticus</i> Hochst.
<i>Cussonia spicata</i> Thunb.
<i>Cyathea dregei</i> Kunze
<i>Cyathea manniana</i> Hook.
<i>Cylicomorpha parviflora</i> Urb.
<i>Diospyros abyssinica</i> (Hiern) F.White
<i>Dodonaea viscosa</i> (L.) Jacq.
<i>Dombeya torrida</i> (J.F.Gmel.) Bamps
<i>Dovyalis macrocalyx</i> (Oliv.) Warb.
<i>Dracaena fragrans</i> (L.) Ker Gawl.
<i>Dracaena steudneri</i> Engl.
<i>Drypetes gerrardii</i> Hutch.
<i>Ehretia cymosa</i> Thonn.
<i>Ekebergia capensis</i> Sparrm.
<i>Embelia schimperii</i> Vatke
<i>Entandrophragma excelsum</i> (Dawe & Sprague) Sprague
<i>Eugenia capensis</i> (Eckl. & Zeyh.) Harv.
<i>Euphorbia abyssinica</i> J.F.Gmel.
<i>Fagaropsis angolensis</i> (Engl.) H.M.Gardner
<i>Ficalhoa laurifolia</i> Hiern
<i>Ficus exasperata</i> Vahl
<i>Ficus natalensis</i> Hochst.
<i>Ficus ovata</i> Vahl
<i>Ficus sur</i> Forssk.
<i>Ficus thonningii</i> Blume
<i>Fleroya rubrostipulata</i> (K.Schum.) Y.F.Deng
<i>Galiniera saxifraga</i> (Hochst.) Bridson
<i>Garcinia buchananii</i> Baker
<i>Grewia arborea</i> (Forssk.) Lam.
<i>Grewia damine</i> Gaertn.
<i>Grewia ferruginea</i> Hochst. ex A.Rich.
<i>Grewia mollis</i> Juss.
<i>Grewia similis</i> K.Schum.
<i>Grewia tembensis</i> Fresen.
<i>Harungana madagascariensis</i> Lam. ex Poir.
<i>Ilex mitis</i> (L.) Radlk.
<i>Kigelia africana</i> subsp. <i>moosa</i> (Sprague) Bidgood & Verdc.
<i>Lecaniodiscus fraxinifolius</i> Baker
<i>Lepidotrichilia volkensii</i> (Gürke) J.-F.Leroy
<i>Macaranga capensis</i> (Baill.) Sim
<i>Maesa lanceolata</i> Forssk.
<i>Manilkara butugi</i> Chiov.
<i>Margaritaria discoidea</i> (Baill.) G.L.Webster

(Continues)

**TABLE 1** (Continued)

<i>Maytenus acuminata</i> (L.f.) Loes.
<i>Maytenus undata</i> (Thunb.) Blakelock
<i>Milicia excelsa</i> (Welw.) C.C.Berg
<i>Mimusops kummel</i> Bruce ex A.DC.
<i>Myrianthus holstii</i> Engl.
<i>Neoboutonia macrocalyx</i> Pax
<i>Newtonia buchananii</i> (Baker) G.C.C.Gilbert & Boutique
<i>Nuxia congesta</i> R.Br. ex Fresen.
<i>Ochna holstii</i> Engl.
<i>Ocotea kenyensis</i> (Chiov.) Robyns & R.Wilczek
<i>Ocotea usambarensis</i> Engl.
<i>Olea capensis</i> L.
<i>Olea europaea</i> L.
<i>Parinari excelsa</i> Sabine
<i>Peddiea africana</i> Harv.
<i>Pittosporum viridiflorum</i> Sims
<i>Plectranthus</i> spp.
<i>Podocarpus falcatus</i> (Thunb.) R. Br. ex Mirb.
<i>Podocarpus henkelii</i> Stapf ex Dallim. & B.D.Jacks.
<i>Podocarpus latifolius</i> (Thunb.) R.Br. ex Mirb.
<i>Polyscias fulva</i> (Hiern) Harms
<i>Pouteria adolfi-friedericii</i> (Engl.) A.Meeuse
<i>Pouteria altissima</i> (A.Chev.) Baehni
<i>Prunus africana</i> (Hook.f.) Kalkman
<i>Psychotria mahonii</i> C.H.Wright
<i>Psychotria orophila</i> E.M.A.Petit
<i>Psydrax parviflora</i> (Afzel.) Bridson
<i>Psydrax schimperiana</i> (A.Rich.) Bridson
<i>Rapanea melanophloeos</i> (L.) Mez
<i>Rhamnus prinoides</i> L'Heri.
<i>Ritchiea albersii</i> Gilg
<i>Rothea myricoides</i> (Hochst.) Steane & Mabb.
<i>Rothmannia urcelliformis</i> (Hiern) Bullock ex Robyns
<i>Rubus apetalus</i> Poir.
<i>Schefflera abyssinica</i> (Hochst. ex A.Rich.) Harms
<i>Schefflera volkensii</i> (Harms) Harms
<i>Scutia myrtina</i> (Burm.f.) Kurz
<i>Senna didymobotrya</i> (Fresen.) H.S.Irwin & Barneby
<i>Shirakiopsis elliptica</i> (Hochst.) Esser
<i>Sinarundinaria alpina</i> (K. Schum.) C.S. Chao & Renvoize
<i>Solanum aculeastrum</i> Dunal
<i>Strombosia scheffleri</i> Engl.
<i>Strychnos mitis</i> S.Moore
<i>Symphonia globulifera</i> L.f.
<i>Synsepalum brevipes</i> (Baker) T.D.Penn.
<i>Syzygium cordatum</i> Hochst. ex Krauss

(Continues)

**TABLE 1** (Continued)

<i>Syzygium guineense</i> (Willd.) DC.
<i>Tabernaemontana pachysiphon</i> Stapf
<i>Tabernaemontana stapfiana</i> Britten
<i>Tarennia graveolens</i> (S. Moore) Bremek.
<i>Teclea nobilis</i> Delile
<i>Trema orientalis</i> (L.) Blume
<i>Trichilia dregeana</i> Sond.
<i>Trilepisium madagascariense</i> DC.
<i>Turraea holstii</i> Gürke
<i>Vachellia drepanolobium</i> (B.Y. Sjöstedt) P.J.H. Hurter
<i>Vachellia hockii</i> (De Wild.) Seigler & Ebinger
<i>Vachellia kirkii</i> (Oliv.) Kyal. & Boatwr.
<i>Vachellia lahai</i> (Steud. & Hochst. ex Benth.) Kyal. & Boatwr.
<i>Vepris dainellii</i> (Pic. Serm.) Mziray
<i>Vepris nobilis</i> (Delile) Mziray
<i>Vernonia auriculifera</i> Hiern
<i>Vernonia myriantha</i> Hook. f.
<i>Vitex fischeri</i> Gürke
<i>Warburgia salutaris</i> (G.Bertol.) Chiov.
<i>Xymalos monospora</i> (Harv.) Baill.
<i>Zanthoxylum gillettii</i> (De Wild.) P.G.Waterman
<i>Zanthoxylum rubescens</i> Planch. ex Hook

Spore stage 0: Spermogonia (consists of receptive hyphae and spermatia)

Spore stage I: Aeciospores in aecia

Spore stage II: Urediniospores in uredia

Spore stage III: Teliospores in telia

Spore stage IV: Basidiospores on basidia

## 2.5 | Automated text mining (ATM) approach

Automated text mining of the biomedical literature has been widely used to recognize entities such as species, proteins, or diseases in the scholarly literature, for example, (Pafilis et al., 2013; Piñero et al., 2015). To our knowledge, this is the first attempt to apply this methodology in the context of plant flora data. Dictionary-based text mining uses a fixed set of identifiers and synonyms that are matched to the contents of scientific articles to identify articles mentioning an entity of interest.

We used text mining to identify species that are mentioned with *C. arabica* in an automated manner. We used the ORGANISMS web resource (Pafilis et al., 2013) to programmatically identify taxa mentioned with *C. arabica* (NCBI Taxonomy ID: 13443) in PubMed abstracts. ORGANISMS recognizes terms from the NCBI Taxonomy via ATM and is updated weekly. We used a version of ORGANISMS downloaded on 9 April 2019. Using a custom Python script, we first identified all PubMed abstracts mentioning *C. arabica* from the ORGANISMS download files. This set of abstracts

**TABLE 2** Medium-ranking list of species considered as potential alternate host(s) of *Hemileia vastatrix*, based on occurrence at the sites of interest in Ethiopia and Kenya, as well as related plant species at the sites of interest in Sri Lanka

Plant species found to occur at the sites of interest in Ethiopia and Kenya <sup>a</sup>	Related plant species found at the sites of interest in Sri Lanka
<i>Acacia abyssinica</i> Hochst. ex Benth.	<i>Acacia decurrens</i> Willd. <i>Acacia leucophloea</i> Willd. <i>Acacia mangium</i> Willd. <i>Acacia melanoxylon</i> R.Br. in W.T.Aiton <i>Acacia planifrons</i> Koenig ex Wight & Arn.
<i>Acacia lahai</i> Steud. & Hochst. ex Benth.	<i>Acacia decurrens</i> Willd. <i>Acacia leucophloea</i> Willd. <i>Acacia mangium</i> Willd. <i>Acacia melanoxylon</i> R.Br. in W.T.Aiton <i>Acacia planifrons</i> Koenig ex Wight & Arn.
<i>Albizia coriaria</i> Oliv.	<i>Albizia falcataria</i> (L.) Fosberg <i>Albizia lebbek</i> sensu auct. <i>Albizia odoratissima</i> (L.f.) Benth. <i>Albizia saman</i> (Jacq.) Merr.
<i>Albizia grandibracteata</i> Taub.	<i>Albizia falcataria</i> (L.) Fosberg <i>Albizia lebbek</i> sensu auct. <i>Albizia odoratissima</i> (L.f.) Benth. <i>Albizia saman</i> (Jacq.) Merr.
<i>Albizia gummifera</i> C.A. Sm.	<i>Albizia falcataria</i> (L.) Fosberg <i>Albizia lebbek</i> sensu auct. <i>Albizia odoratissima</i> (L.f.) Benth. <i>Albizia saman</i> (Jacq.) Merr.
<i>Albizia schimperiana</i> Oliv.	<i>Albizia falcataria</i> (L.) Fosberg <i>Albizia lebbek</i> sensu auct. <i>Albizia odoratissima</i> (L.f.) Benth. <i>Albizia saman</i> (Jacq.) Merr.
<i>Allophylus abyssinicus</i> (Hochst) Radlk	<i>Allophylus cobbe</i> (L.) Raeusch.
<i>Apodytes dimidiata</i> E. Mey. ex Arn.	<i>Apodytes dimidiata</i> E.Mey. ex Arn.
<i>Cassipourea malosana</i> Alston	<i>Cassipourea</i> spp.
<i>Cassipourea ruwensensis</i> Alston	<i>Cassipourea</i> spp.
<i>Croton macrostachyus</i> Hochst. ex Delile	<i>Croton moonii</i> Thwaites
<i>Croton megalocarpus</i> Hutch.	<i>Croton moonii</i> Thwaites
<i>Croton sylvaticus</i> Hochst.	<i>Croton moonii</i> Thwaites

**TABLE 2** (Continued)

Plant species found to occur at the sites of interest in Ethiopia and Kenya <sup>a</sup>	Related plant species found at the sites of interest in Sri Lanka
<i>Diospyros abyssinica</i> (Hiern) F.White	<i>Diospyros ebenum</i> J.Koenig ex Retz. <i>Diospyros malabarica</i> (Desr.) Kostel. <i>Diospyros melanoxylon</i> Roxb. <i>Diospyros oocarpa</i> Thwaites <i>Diospyros ovalifolia</i> Wight
<i>Macaranga capensis</i> (Baill.) Sim.	<i>Macaranga indica</i> Wight <i>Macaranga peltata</i> (Roxb.) Müll.Arg.
<i>Ochna holstii</i> Engl.	<i>Ochna rufescens</i> Thwaites <i>Ochna lanceolata</i> Kuntze
<i>Psychotria mahonii</i> Hook. f.	<i>Psychotria moonii</i> Hook.f.
<i>Solanum aculeastrum</i> Dunal	<i>Solanum aculeastrum</i> Dunal (reported in Southern India only)
<i>Strychnos mitis</i> S. Moore	<i>Strychnos nux-vomica</i> L.
<i>Syzygium cordatum</i> Hochst.	<i>Syzygium aqueum</i> (Burm.f.) Alston <i>Syzygium assimile</i> Thwaites <i>Syzygium cumini</i> (L.) Skeels <i>Syzygium gardneri</i> Thwaites <i>Syzygium jambos</i> (L.) Alston <i>Syzygium lewisii</i> Alston <i>Syzygium makul</i> Gaertn. <i>Syzygium neesianum</i> Arn. <i>Syzygium operculatum</i> (Roxb.) Nied. <i>Syzygium rubicundum</i> Wight & Arn. <i>Syzygium umbrosum</i> Thwaites <i>Syzygium zeylanicum</i> (L.) DC.
<i>Syzygium guineense</i> DC.	<i>Syzygium aqueum</i> (Burm.f.) Alston <i>Syzygium assimile</i> Thwaites <i>Syzygium cumini</i> (L.) Skeels <i>Syzygium gardneri</i> Thwaites <i>Syzygium jambos</i> (L.) Alston <i>Syzygium lewisii</i> Alston <i>Syzygium makul</i> Gaertn. <i>Syzygium neesianum</i> Arn. <i>Syzygium operculatum</i> (Roxb.) Nied. <i>Syzygium rubicundum</i> Wight & Arn. <i>Syzygium umbrosum</i> Thwaites <i>Syzygium zeylanicum</i> (L.) DC.
<i>Vitex fischeri</i> Gürke	<i>Vitex altissima</i> L.f. <i>Vitex leucoxydon</i> L.f.

<sup>a</sup>Plant species listed were classified as either “characteristic” or “present” at the sites of interest in Ethiopia and Kenya according to the PNV maps. Plants are listed in alphabetical order. Plant names and authors verified by The Taxonomic Name Resolution Service.

(Continues)



was then intersected with the set of abstracts mentioning each taxon to determine the number of abstracts mentioning *C. arabica* and the respective taxon. Since the ORGANISMS web resources uses fully automated pattern matching to identify taxa from the NCBI Taxonomy in PubMed abstracts, the mentions contained in ORGANISMS may contain both false-positive matches (the taxon is falsely recognized in the abstract) and false-negative matches (the taxon is falsely overlooked in the abstract). Any plant species that are found in the NCBI taxonomy lineage for *C. arabica* are automatically listed as a mentioned item in the ATM results (<https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?xml:id=13443>).

## 2.6 | Comparison of the HAHR and the ATM methods

A manual cross-referencing approach was applied to the list of species generated by the ATM method by using the “find” function in MS Excel for all of the plant species listed in the HAHR method (Tables 1–3). The ranking established in the HAHR was also applied to the ATM (Table 4). The abstracts listed by the ATM method were reviewed to assess whether both *C. arabica* and the potential plant host species were evident in the publication and how many times they were mentioned (Table 4). The percentage difference in overlap of potential host species was calculated as a way to compare the HAHR and ATM methods output.

## 3 | RESULTS AND DISCUSSION

The HAHR method indicated 158 plant species as potential alternate hosts of *Hemileia vastatrix*, while the ATM method listed over 2,179 species (although some duplication was found). There were 19 plant species, which overlapped both methods (Table 4). Indicating that 12% of the HAHR findings were corroborated by the ATM method. The low overlap percentage reflects the variation in the two methods.

The HAHR method produced a low-ranking short list (Table 1), a medium-ranking list (Table 2), and a high-ranking list (Table 3) of plant species. Table 1 comprises 158 plant species, which were found at the sites of interest in Ethiopia and Kenya in flora mapping studies concomitantly with *Coffea* species. The absence of common species between the sites of interest in Ethiopia, Kenya, and Sri Lanka (Table 2) could be seen as an indication that there would be more than one possible alternate host species of CLR. High rates of endemism caused by the long period of isolation of the island (Gunatilleke & Gunatilleke, 1990) mean that over a quarter of the native species present in Sri Lanka are considered unique to the country (Ashton et al., 1997). However, the Deccan-Gondwana ancestry of the country means there may have been an early contribution to the natural plant communities of Sri Lanka's west coast from the African continent (Ashton et al., 1997). This may explain why there are similarities at the genus level at the sites of interest in Ethiopia, Kenya, and Sri Lanka (Table 2).

Table 3 shows 22 high-ranking, potential alternate hosts of *H. vastatrix*, belonging to 10 different families (Cornaceae, Euphorbiaceae, Hypericaceae, Lamiaceae, Moraceae, Phyllanthaceae, Podocarpaceae, Rhamnaceae, Rosaceae, and Rubiaceae). This list includes conifers and angiosperm dicots such as succulents, herbs, shrubs, and trees. Table 3 has been divided into three additional categories of ranking. Category 1 ranking implies that the plant species or genera have been observed at all the sites of interest and has a reported susceptibility to *Hemileia* spp. or other rust pathogens. Category 2 ranking implies that the plant species or genera have been observed at two sites of interest and have a known susceptibility to a *Hemileia* species. Category 3 ranking implies that the plant species or genera have been observed at two sites of interest and have a known susceptibility to other rust pathogens. The highest-ranking plant species found were *Psychotria mahonii*, *Rubus apetalus*, and *Rhamnus prinoides*. All three species were present at the sites of interest in Ethiopia and Kenya as well as having a presence at the genus level in Sri Lanka. Furthermore, these plant species are known hosts of *Hemileia holstii* (MyCoPortal, 2018), *Kuehneola uredinis* (Van Reenen, 1995), and *Puccinia coronata* (Nazareno et al., 2017), respectively.

Overall, minimal botanical data were found for *P. mahonii*. This plant species is stated as being either a tree (up to 15 m) or shrub (up to 5 m) and native to Burundi, Cameroon, Gabon, Kenya, Malawi, Rwanda, Sudan, Tanzania, Uganda, Zambia, Zaïre, and Zimbabwe (Flora of Tropical East Africa, 2018). Given that *Psychotria* spp. are native and widespread across the tropics and subtropical regions (including Sri Lanka), the likelihood of interaction with *H. vastatrix* is high and therefore makes this species a high-ranking candidate as an alternate host of the fungus.

*Rubus apetalus* has been documented as a host of the rust fungus *Kuehneola uredines* (casual pathogen of blackberry cane and leaf rust) in the uredinial stage (Van Reenen, 1995). Interestingly, the telial, spermogonial, and aecial states of *K. uredinis* have also been observed on *Rubus penetrans* (Gardner & Hodges, 1983), thus making this plant species a possible aecial host for other rust fungi. Similarly, *R. prinoides* is the alternate host of *Puccinia coronata* (causal pathogen of crown rust in cultivated and wild oat) (Nazareno et al., 2017). *Rhamnus* and *Rubus* spp. are known for occurring pervasively across a wide range of habitats including mountain forests, especially in clearings and along edges; along watercourses; in riverine forests; in margins of evergreen forests; in secondary mountain evergreen forests or bushes; on mountain slopes; and in grasslands (<http://ecocrop.fao.org>; Ruffo, Birnie, & Tengnäs, 2002). Both the genera *Rhamnus* and *Rubus* are widespread throughout Asia, Africa, and the Americas, which are today coffee leaf rust hotspots.

The ATM method recovered 700 articles, which mentioned *C. arabica* and 2,179 taxa were mentioned at least once with *C. arabica*. These taxa included plants, animals, fungi, and bacteria. The 2,177 taxa were sorted by the number of PubMed abstracts that mentioned them with *C. arabica*. The relevant plant genera derived from the ATM and manual analysis are shown in Table 4. Several genera

**TABLE 3** High-ranking list of species considered as potential alternate host(s) of *Hemileia vastatrix*, based on the hypothetical alternate host ranking (HAHR)

Plant species <sup>a</sup>	The sites of interest	Known interactions with <i>Hemileia</i> spp. or other rust pathogens
Category 1 ranking		
<i>Psychotria mahonii</i> C.H.Wright	E, K & S <sup>b</sup>	Genus is susceptible to <i>Hemileia holstii</i> P. Syd. & Syd. (MyCoPortal, 2018)
<i>Rhamnus prinoides</i> L'Heri	E, K & S <sup>b</sup>	Genus is susceptible to rust fungus: <i>Puccinia coronata</i> Peterson reported on <i>Rhamnus</i> spp. in the spermogonial (O) and aecial rust stages (I) (Arthur, 1934 pp. 152; Nazareno et al., 2017)
<i>Rubus apetalus</i> Poir.	E, K & S <sup>b</sup>	Species and genus are susceptible to rust fungus: <i>Kuehneola uredinis</i> (Link) Arthur in either the uredinial (II) and telial (III) rust stages (MyCoPortal, 2018; Van Reenen, 1995)
Category 2 ranking		
<i>Antidesma venosum</i> E.Mey. ex Tul.	E, K	Genus is susceptible to <i>Hemileia antidesmae</i> P. Syd. & Syd. (MyCoPortal, 2018)
<i>Canthium keniense</i> Bullock	E, S <sup>b</sup>	Genus is susceptible to <i>Hemileia canthi</i> (Berk. & Broome) in the spermogonial (O) and aecial rust stages (I) (MyCoPortal, 2018)
<i>Canthium lactescens</i> Hiern	E, S <sup>b</sup>	Genus is susceptible to <i>Hemileia canthi</i> (Berk. & Broome) in the spermogonial (O) and aecial rust stages (I) (MyCoPortal, 2018)
<i>Canthium oligocarpum</i> Hiern	E, S <sup>b</sup>	Genus is susceptible to <i>Hemileia canthi</i> (Berk. & Broome) in the spermogonial (O) and aecial rust stages (I) (MyCoPortal, 2018)
<i>Clerodendrum myricoides</i> (Hochst.) R.Br. ex Vatke	E, S <sup>b</sup>	Genus is susceptible to <i>Hemileia scholzii</i> Syd. & P. Syd. (MyCoPortal, 2018)
<i>Harungana madagascariensis</i> Poir.	E, K	Species is susceptible to <i>Hemileia harunganae</i> Cummins (MyCoPortal, 2018)
<i>Vangueria apiculata</i> K. Schum.	E, K	Genus is susceptible to <i>Hemileia thomasi</i> Thirum. & Naras. and <i>Hemileia woodii</i> Kalchbr. & Cooke
Category 3 ranking		
<i>Alchemilla adolfi-friederici</i> Engl.	E, K	Genus is susceptible to <i>Trachyspora alchemillae</i> (Pers.) Fuckel in the uredinial rust stage (II) (Helfer, 2005; MyCoPortal, 2018) and reported susceptibility to decay by basidiomycete fungi (Desalegn, 2013)
<i>Cornus volkensii</i> Harms	E, K	Genus is susceptible to rust fungus: <i>Puccinia porphyrogenita</i> M.A. Curtis reported on <i>Cornus canadensis</i> in either the uredinial (II) or telial (III) rust stages (Arthur, 1934 pp. 251; MyCoPortal, 2018)
<i>Croton dichogamus</i> Pax	E, S <sup>b</sup>	Genus is susceptible to rust fungus: <i>Bubakia crotonis</i> Arthur reported on <i>Croton argyranthemus</i> , <i>Croton californicus</i> , <i>Croton capitatus</i> , <i>Croton engelmannii</i> , <i>Croton monanthogynus</i> , <i>Croton punctatus</i> , and <i>Croton texensis</i> in either the uredinial (II) or telial (III) rust stages (Arthur, 1934 pp. 60; MyCoPortal, 2018)
<i>Croton macrostachyus</i> Hochst. ex Delile	E, S <sup>b</sup>	Genus is susceptible to rust fungus: <i>Bubakia crotonis</i> Arthur reported on <i>Croton argyranthemus</i> , <i>Croton californicus</i> , <i>Croton capitatus</i> , <i>Croton engelmannii</i> , <i>Croton monanthogynus</i> , <i>Croton punctatus</i> , and <i>Croton texensis</i> in either the uredinial (II) or telial (III) rust stages (Arthur, 1934 pp. 60; MyCoPortal, 2018)
<i>Croton sylvaticus</i> Hochst.	E, S <sup>b</sup>	Genus is susceptible to rust fungus: <i>Bubakia crotonis</i> Arthur reported on <i>Croton argyranthemus</i> , <i>Croton californicus</i> , <i>Croton capitatus</i> , <i>Croton engelmannii</i> , <i>Croton monanthogynus</i> , <i>Croton punctatus</i> , and <i>Croton texensis</i> in either the uredinial (II) or telial (III) rust stages (Arthur, 1934 pp. 60; MyCoPortal, 2018)
<i>Euphorbia abyssinica</i> J.F.Gmel.	E, K	Genus is susceptible to rust fungus: <i>Melampsora euphorbiae</i> (Ficinus & C. Schub.) Castagne is found on a wide range of <i>Euphorbia</i> spp. in either I, II, III, or O rust stages (Wilson & Henderson, 1966 pp. 68 and Arthur, 1934 pp.309; MyCoPortal, 2018)
<i>Ficus ovata</i> Vahl	E, K	Genus is susceptible to rust fungus: <i>Physopella fici</i> (Castagne) Arthur reported on <i>Ficus carica</i> in either the uredinial (II) or telial (III) rust stages (Arthur, 1934 pp. 61; MyCoPortal, 2018)
<i>Ficus sur</i> Forssk.	E, K	Genus is susceptible to rust fungus: <i>Physopella fici</i> (Castagne) Arthur reported on <i>Ficus carica</i> in either the uredinial (II) or telial (III) rust stages (Arthur, 1934 pp. 61; MyCoPortal, 2018)
<i>Ficus thonningii</i> Blume	E, K	Genus is susceptible to rust fungus: <i>Physopella fici</i> (Castagne) Arthur reported on <i>Ficus carica</i> in either the uredinial (II) or telial (III) rust stages (Arthur, 1934 pp. 61; MyCoPortal, 2018)

<sup>a</sup>The plant species or the genera were found to occur at the sites of interest in Ethiopia, Kenya, and/or Sri Lanka and have a known susceptibility to *Hemileia* spp. or other rust pathogens. Plants are listed in a prioritized category ranking order. Category 1 ranking implies that the plant species or genera have been observed at all the sites of interest and have a reported susceptibility to *Hemileia* spp. or other rust pathogens. Category 2 ranking implies that the plant species or genera have been observed at two sites of interest and have a known susceptibility to a *Hemileia* species. Category 3 ranking implies that the plant species or genera have been observed at two sites of interest and have a known susceptibility to other rust pathogens. Plant names and authors verified by The Taxonomic Name Resolution Service. Fungi names and authors verified by Species Fungorum. The sites of interest are stated as E (Ethiopia), K (Kenya), and S (Sri Lanka).

<sup>b</sup>Only present at genus level.

have been comentioned with *C. arabica* in the PubMed databases. The cross-referenced results by the two methods suggest that plant genera of interest are *Rubus*, *Croton*, and *Euphorbia*. *Rubus* is of high interest as it is the only commonly found genus in both the category one ranking of the HAHR and with the ATM method. The genus *Solanum* was very often comentioned with *C. arabica* (49 abstracts) according to the ATM method. However, it was only listed on the medium-ranking list when using the HAHR method. This shows the variation between the two methods as the HAHR is based on published flora data showing co-occurrence at the sites of interest, whereas the ATM shows search results of *C. arabica* plus any other comentioned species independent of the geography or nature of the published study.

Both the methods used in this hypothesis paper are limited by the published literature and databases, which were used as the “data pool” for each of the analyses. The HAHR method relies on published flora data produced in English and is region-specific based on the decision tree (Figure 2) making it a targeted method in the context of this study. On the other hand, the ATM method is restricted to the PubMed databases and includes all species (animals, bacteria, fungi, and plants) comentioned with *C. arabica*, making it more comprehensive than the HAHR method. However, this also leads to superfluous data retrieval, which needed to be manually filtered.

**TABLE 4** Cross-reference of high-, medium-, and low-ranking species with the automated text mining method (ATM)

Plant genus	Number of abstracts comentioned with <i>C. arabica</i>
High-ranking list	
<i>Rubus</i> spp.	5
<i>Croton</i> spp.	3
<i>Euphorbia</i> spp.	8
Medium-ranking list	
<i>Acacia</i> spp.	8
<i>Solanum</i> spp.	49
<i>Syzygium</i> spp.	6
Low-ranking list	
<i>Vachellia</i> spp.	2
<i>Synsepalum</i> spp.	2
<i>Prunus</i> spp.	11
<i>Pittosporum</i> spp.	2
<i>Olea</i> spp.	5
<i>Ocotea</i> spp.	1
<i>Maytenus</i> spp.	2
<i>Garcinia</i> spp.	3
<i>Cyathea</i> spp.	1
<i>Cornus</i> spp.	3
<i>Cordia</i> spp.	1
<i>Catha</i> spp.	2
<i>Alchornea</i> spp.	2

Furthermore, the ATM method may yield false-positive hits due to the species names listed in the NCBI taxonomy, that are falsely recognized in the analyzed articles. Again, manual filtering avoided these false-positive results to be included in the results. Given the incorporation of plant species geography into the HAHR method, it is recommended that the findings from this method be prioritized over the ATM method.

A corroboration of the hypothesis raised in this study would be morphological and molecular examinations of historical plant leaf samples from herbaria collections. Plant species from the ranked listings (Tables 1–3), which were collected during epidemic periods, may exhibit symptoms of CLR, which can help to verify the alternate host. This approach has led to the recent revision of the history and geographical range of *Colletotrichum acutatum* species (Sundelin et al., 2015), as well as the sequencing of a unique genotype of *Phytophthora infestans* (HERB-1), which is now accepted as the causal virulent race which lead to the 19th century potato late blight epidemic (Yoshida et al., 2013).

Based on the HAHR and ATM methods presented here, it is the hope that the alternate host(s) of *H. vastatrix* will be conclusively identified. Priority for future exploratory studies in this context should be offered to plant species *P. mahonii*, *R. apetalus*, and *R. prinoides*, as indicated by the HAHR method. Further exploration of the genera *Croton*, *Euphorbia*, and *Rubus*, as indicated by the cross-referencing with the ATM method, is also of interest. The ranked plant species lists presented here may be used as a guide to search for signs of spermogonia and aecia of *H. vastatrix* under natural infection conditions or on historical specimens, from the areas we have presented as the sites of origin, discovery or first reported outbreak of CLR. By narrowing down the possible alternate host of *H. vastatrix*, it is our hope to help solve the mystery that has been perplexing the plant pathology community for more than 150 years.

## ACKNOWLEDGMENTS

We gratefully acknowledge collaboration with the BREEDCAFS project and thank Jacques Avelino for his expert input relating to *H. vastatrix* and his suggestions for possible alternate plant host candidates. We also thank Benjamin and Charles Robotham for their review of the manuscript and helpful annotations.

## CONFLICT OF INTEREST

The authors declare that they have no competing interests.

## AUTHOR CONTRIBUTIONS

AK was responsible for the HAHR analysis and manual ATM analysis and for developing the draft. AR and JPBL contributed to the initial plant species pool development and flora literature review. HJLJ and BJ contributed to the literature review concerning the pathogen biology and history. AJ was responsible for the adaptation to the ATM method. All authors commented on the manuscript and approved

the final version. None of the authors has any competing interests in the manuscript.

## ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable

## CONSENT FOR PUBLICATION

Not applicable

## DATA AVAILABILITY STATEMENT

Databases generated or analyzed during this study are included in this published article [and its supplementary information files].

## ORCID

Athina Koutouleas  <https://orcid.org/0000-0001-6113-6313>

Hans Jørgen Lyngs Jørgensen  <https://orcid.org/0000-0002-5902-6550>

Birgit Jensen  <https://orcid.org/0000-0003-2476-805X>

Jens-Peter Barnekow Lillesø  <https://orcid.org/0000-0003-1418-3496>

Alexander Junge  <https://orcid.org/0000-0002-2410-9671>

Anders Ræbild  <https://orcid.org/0000-0002-8793-5663>

## REFERENCES

- Arneson, P. A. (2000). Coffee rust. *The Plant Health Instructor* [Internet] retrieved from: <https://www.apsnet.org/edcenter/disandpath/fungalbasidio/pdlessons/Pages/CoffeeRust.aspx>
- Arthur, J. C. (1934). *Manual of the Rusts in United States and Canada* (p. 438). Lafayette, IN: Purdue Research Foundation. Incl. Supplement to Arthur's Manual of the Rusts in United States and Canada by G. B. Cummins. Hafner Publishing Company. New York, 1962.
- Ashton, M. S., & Gunatilleke, S. (1987). New light on the plant geography of Ceylon. I. Historical plant geography. *Journal of Biogeography*, *14*, 249–285.
- Ashton, M. S., Gunatilleke, S., De Zoysa, N., Dassanayake, M. D., Gunatilleke, N., & Wijesundera, S. (1997). *A field guide to the common trees and shrubs of Sri Lanka* (p. 430). Colombo, Sri Lanka: WHT Publications.
- Berkeley, M. J., & Broome, C. E. (1869). *Hemileia vastatrix*. *The Gardeners' Chronicle and Agricultural Gazette*, *6*, 1157.
- Brown, J. K., & Hovmöller, M. S. (2002). Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease. *Science*, *297*, 537–541. <https://doi.org/10.1126/science.1072678>
- Burdon, J. J., & Chilvers, G. A. (1982). Host density as a factor in plant disease ecology. *Annual Review of Phytopathology*, *20*, 143–166. <https://doi.org/10.1146/annurev.py.20.090182.001043>
- Carvalho, C. R., Fernandes, R. C., Carvalho, G. M. A., Barreto, R. W., & Evans, H. C. (2011). Cryptosexuality and the genetic diversity paradox in coffee rust, *Hemileia vastatrix*. *PLoS ONE*, *6*(11), e26387. <https://doi.org/10.1371/journal.pone.0026387>
- Coutinho, T. A., Rijkenberg, F. H. J., & Van Asch, M. A. J. (1995). Teliospores of *Hemileia vastatrix*. *Mycological Research*, *99*, 932–934. [https://doi.org/10.1016/S0953-7562\(09\)80751-X](https://doi.org/10.1016/S0953-7562(09)80751-X)
- Desalegn, G. (2013). Protection measures of *Juniperus procera* Hochst. ex Endl. (Cupressaceae) and *Eucalyptus deglupta* Blume (Myrtaceae) sawn timbers against subterranean termites and fungal attack at Miesso, eastern Ethiopia. *Ethiopian Journal of Biological Sciences*, *12*(2), 187–210.
- FAO – Food and Agriculture Organization of the UN [Internet]. *Ecocrop database*. Retrieved from [www.ecocrop.fao.org](http://www.ecocrop.fao.org)
- Ferguson, A. M., & Ferguson, J. (1882). A monthly record of information for planters of coffee, cocoa, cinchona, India-rubber, sugar, tobacco, cardamons, palms, rice and other products suited for cultivation in the tropics. *The Tropical Agriculturist*, *1*, 603.
- Ferreira, S. A., & Boley, R. A. (1991). *Hemileia vastatrix*. Crop Knowledge Master. Retrieved from [www.extento.hawaii.edu/kbase/crop/type/h\\_vasta.htm](http://www.extento.hawaii.edu/kbase/crop/type/h_vasta.htm)
- Flora of Tropical East Africa (2018). *Kew Royal Botanical Gardens*. Retrieved from <http://www1.kew.org/science/directory/projects/FloraTropEAfrica.html>
- Gardner, D. E., & Hodges, C. S. (1983). Leaf rust caused by *Kuehneola uredinis* on native and non-native *Rubus* species in Hawaii. *Plant Disease*, *67*, 962–963.
- Gole, T. W. (2003). Vegetation of the Yayu forest in SW Ethiopia: Impacts of human use and implications for in situ conservation of wild *Coffea arabica* L. populations. *Ecology and Development Series*, *10*, 2–8.
- Gopalkrishnan, K. S. (1951). Notes on the morphology of the genus *Hemileia*. *Mycologia*, *43*, 271–283. <https://doi.org/10.1080/00275514.1951.12024128>
- Gunatilleke, I. A. U. N., & Gunatilleke, C. V. S. (1990). Distribution of floristic richness and its conservation in Sri Lanka. *Conservation Biology*, *4*, 21–31.
- Helfer, S. (2005). Overview of the rust fungi (Uredinales) occurring on Rosaceae in Europe. *Nova Hedwigia*, *81*, 325–370. <https://doi.org/10.1127/0029-5035/2005/0081-0325>
- Hiratsuka, Y., & Sato, S. (1982). Morphology and taxonomy of rust fungi. In K. J. Scott, & A. K. Chakravorty (Eds.), *The rust fungi* (pp. 1–36). New York, NY: Academic Press.
- Jin, Y., Szabo, L. J., & Carson, M. (2010). Century-old mystery of *Puccinia striiformis* life history solved with the identification of *Berberis* as an alternate host. *Phytopathology*, *100*, 432–435.
- Kaitera, J., Hiltunen, R., & Hantula, J. (2017). *Nasa*, *Nemesia* and *Euphrasia*: New alternate hosts of *Cronartium* spp. *Forest Pathology*, *47*(2), e12306.
- Kelbessa, E., & Soromessa, T. (2008). Interfaces of regeneration, structure, diversity and uses of some plant species in Bonga Forest: A reservoir for wild coffee gene pool. *SINET: Ethiopian Journal of Science*, *31*, 121–134.
- Kolmer, J. A., Ordóñez, M. E., & Groth, J. V. (2009). *The rust fungi*. eLS. Retrieved from [https://ars.usda.gov/ARSUserFiles/3094/rust\\_fungi.pdf](https://ars.usda.gov/ARSUserFiles/3094/rust_fungi.pdf)
- Kumar, A., Sreedharan, S. P., Shetty, N. P., & Parvatam, G. (2016). Developing sustainable disease resistance in coffee. In D. B. Collinge (Ed.), *Plant pathogen resistance biotechnology* (1st ed., pp. 217–243). Copenhagen, Denmark: John & Wiley Sons.
- Lashermes, P., Bertrand, B., & Ettienne, H. (2009). Breeding coffee (*Coffea arabica*) for sustainable production. In: S. M. Jain, & P. M. Priyadarshan (Eds.), *Breeding plantation tree crops: Tropical species* (pp. 525–543). New York, NY: Springer.
- McCook, S. (2006). Global rust belt: *Hemileia vastatrix* and the ecological integration of world coffee production since 1850. *Journal of Global History*, *1*, 177–195.
- McDonald, G. I., Richardson, B. A., Zambino, P. J., Klopfenstein, N. B., & Kim, M. S. (2006). *Pedicularis* and *Castilleja* are natural hosts of *Cronartium ribicola* in North America: A first report. *Forest Pathology*, *36*, 73–82. <https://doi.org/10.1111/j.1439-0329.2006.00432.x>
- Meyer, F. G. (1965). Notes on wild *Coffea arabica* from Southwestern Ethiopia, with some historical considerations. *Economic Botany*, *19*, 136–151. <https://doi.org/10.1007/BF02862825>

- MyCoPortal (2018). [Internet]. Retrieved from <http://mycoportal.org/portal/index.php>
- Nazareno, E. S., Li, F., Smith, M., Park, R. F., Kianian, S. F., & Figueroa, M. (2017). *Puccinia coronata* f.sp. *avenae*: A threat to global oat production. *Molecular Plant Pathology*, 19, 1047–1060.
- Nune, S. (2008). *Flora biodiversity assessment in Bonga, Boginda and Mankira Forest, Kafa, Ethiopia*. Addis Ababa, Ethiopia: Ethiopian Wildlife and Natural History Society. Retrieved from <http://www.kafa-biosphere.com/assets/content-documents/KafaFloral-Survey-Final-Report.pdf>
- Nutman, F. J., Roberts, F. M., & Bock, K. R. (1960). Method of uredospore dispersal of the coffee leaf-rust fungus, *Hemileia vastatrix*. *Transactions of the British Mycological Society*, 43, 509–515.
- Pafilis, E., Frankild, S. P., Fanini, L., Faulwetter, S., Pavlouidi, C., Vasileiadou, A., ... Jensen, L. J. (2013). The SPECIES and ORGANISMS resources for fast and accurate identification of taxonomic names in text. *PLoS ONE*, 8(6), e65390. <https://doi.org/10.1371/journal.pone.0065390>
- Petersen, R. H. (1974). The rust fungus life cycle. *The Botanical Review*, 40, 453–513.
- Piñero, J., Queralt-Rosinach, N., Bravo, À., Deu-Pons, J., Bauer-Mehren, A., Baron, M., ... Furlong, L. I. (2015). *DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes*. Database, 2015, bav028.
- Rodrigues, C. J. Jr (1990). Coffee rusts: History, taxonomy, morphology, distribution and host resistance. *Fitopatol Brasileira*, 15, 5–9.
- Rodriguez-Algaba, J., Walter, S., Sørensen, C. K., Hovmøller, M. S., & Justesen, A. F. (2014). Sexual structures and recombination of the wheat rust fungus *Puccinia striiformis* on *Berberis vulgaris*. *Fungal Genetics and Biology*, 70, 77–85. <https://doi.org/10.1016/j.fgb.2014.07.005>
- Ruffo, C. K., Birnie, A., & Tengnäs, B. (2002). *Edible wild plants of Tanzania. Technical handbook* (Vol. 27, p. 766). Nairobi, Kenya: Regional Land Management Unit/Sida. Retrieved from <http://www.worldagroforestry.org/downloads/Publications/PDFS/B11913.pdf>
- Schmitt, C. B. (2006). Montane rainforest with wild *Coffea arabica* in the Bonga region (SW Ethiopia): plant diversity, wild coffee management and implications for conservation. In P. L. G. Vlek (Ed.), *Ecology and development systems* (Vol. 47, pp. 1–28). Göttingen, Germany: Cuvillier Verlag.
- Schumann, G. L., & Leonard, K. J. (2000). Stem rust of wheat (black rust). *The Plant Health Instructor*. <https://doi.org/10.1094/PHI-I-2000-0721-01>. Updated 2011.
- Senbeta, F., & Denich, M. (2006). Effects of wild coffee management on species diversity in the Afromontane rainforests of Ethiopia. *Forest Ecology and Management*, 232, 68–74. <https://doi.org/10.1016/j.foreco.2006.05.064>
- Shattock, R. C., & Preece, T. F. (2000). Tranzschel revisited: Modern studies of the relatedness of different rust fungi confirm his law. *Mycologist*, 14, 113–117. [https://doi.org/10.1016/S0269-915X\(00\)80086-5](https://doi.org/10.1016/S0269-915X(00)80086-5)
- Silva, D. N., Várzea, V., Paulo, O. S., & Batista, D. (2018). Population genomic footprints of host adaptation, introgression and recombination in coffee leaf rust. *Molecular Plant Pathology*, 19, 1742–1753. <https://doi.org/10.1111/mpp.12657>
- Species Fungorum [Internet]. Retrieved from <http://www.speciesfungorum.org/names/names.asp>
- Sundelin, T., Strömeng, G. M., Gjærum, H. B., Amby, D. B., Ørstad, K., Jensen, B., ... Stensvand, A. (2015). A revision of the history of the *Colletotrichum acutatum* species complex in the Nordic countries based on herbarium specimens. *FEMS Microbiology Letters*, 362(16), p.fnv130.
- Tadesse, M., & Nigatu, L. (1996). An ecological and ethnobotanical study of wild or spontaneous coffee, *Coffea arabica* in Ethiopia. In L. J. G. van der Maesen, X. M. van der Burgt, & J. M. van Medenbach de Rooy (Eds.), *The biodiversity of African plants* (pp. 277–294). Wageningen, The Netherlands: Springer.
- Talhinhas, P., Batista, D., Diniz, I., Vieira, A., Silva, D. N., Loureiro, A., ... Várzea, V. (2017). The coffee leaf rust pathogen *Hemileia vastatrix*: One and a half centuries around the tropics. *Molecular Plant Pathology*, 18, 1039–1051.
- The Global Biodiversity Information Facility [Internet]. Retrieved from <https://www.gbif.org/>
- The Plant List [Internet]. Version 1.1. Retrieved from <http://www.theplantlist.org/>
- The Taxonomic Name Resolution Service [Internet]. *iPlant collaborative*. Version 4.0. Retrieved from <http://tnrs.iplantcollaborative.org>
- Toniutti, L., Breitler, J. C., Etienne, H., Campa, C., Doubeau, S., Urban, L., ... Bertrand, B. (2017). Influence of environmental conditions and genetic background of *Arabica coffee* (*C. arabica* L.) on leaf rust (*Hemileia vastatrix*) pathogenesis. *Frontiers. Plant Science*, 8, 2025. <https://doi.org/10.3389/fpls.2017.02025>
- van Breugel, P., Kindt, R., Lillesø, J. P. B., Bingham, M., Demissew, S., Dudley, C., ... Graudal, L. (2015). *Potential natural vegetation map of Eastern Africa (Burundi, Ethiopia, Kenya, Malawi, Rwanda, Tanzania, Uganda and Zambia)*. Version 2.0. Forest and Landscape (Denmark) and World Agroforestry Centre (ICRAF). Retrieved from <http://vegetationmap4africa.org>
- Van Reenen, M. (1995). An annotated list of Urediniomycetes (rust fungi) from South Africa 1: Melampsoraceae and Pucciniaceae, excluding *Puccinia* and *Uromyces*. *Bothalia*, 25, 173–181.
- Waller, J. M. (1982). Coffeerust – Epidemiology and control. *Crop Protection*, 1, 385–404. [https://doi.org/10.1016/0261-2194\(82\)90022-9](https://doi.org/10.1016/0261-2194(82)90022-9)
- Ward, H. M. (1882). Researches on the life-history of *Hemileia vastatrix*, the fungus of the “Coffee-leaf-disease”. *Journal of the Linnean Society, Botany*, 19, 299–335.
- Wilson, M., & Henderson, D. M. (1966). *British rust fungi* (p. 68). Cambridge, UK: Cambridge University Press.
- Yoshida, K., Schuenemann, V. J., Cano, L. M., Pais, M., Mishra, B., Sharma, R., ... Burbano, H. A. (2013). The rise and fall of the *Phytophthora infestans* lineage that triggered the Irish potato famine. *eLife*, 2, e00731. <https://doi.org/10.7554/eLife.00731>

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Koutouleas A, Jørgen Lyngs Jørgensen H, Jensen B, Lillesø J-PB, Junge A, Ræbild A. On the hunt for the alternate host of *Hemileia vastatrix*. *Ecol Evol*. 2019;9:13619–13631. <https://doi.org/10.1002/ece3.5755>