



# Genomic resources of *Colletotrichum* fungi: development and application

Ayako Tsushima<sup>1</sup> · Ken Shirasu<sup>2,3</sup>

Received: 19 July 2022 / Accepted: 5 September 2022 / Published online: 17 October 2022  
© The Author(s) 2022

## Abstract

Anthraco­nose caused by *Colletotrichum* spp. is an economically important disease of many plants, including grain, vegetable, and fruit crops. Next-generation sequencing technologies have led to a dramatic growth in the size and availability of genomic data in public repositories. Beginning with genome sequencing projects of *C. higginsianum* and *C. gramini­cola*, many *Colletotrichum* spp. genomes have been sequenced due to their scientific and agricultural importance. Today, we can access more than a hundred genome assemblies of *Colletotrichum* spp. Utilizing those abundant genomic datasets would enable a better understanding of adaptation mechanisms of *Colletotrichum* spp. at the genomic level, which could help to control this important group of pathogens. In this review, we outline the development and application of genomic resources of *Colletotrichum* spp. with a focus on the benefits of genomic data-driven studies, including reverse-genetics, a range of comparative genomic analyses, species identification, taxonomy, and diagnosis, while describing the potential pitfalls of genome analysis. Further, we discuss future research directions that could allow a more comprehensive understanding of genomic diversity within the genus *Colletotrichum*.

**Keywords** *Colletotrichum* · Anthracnose · Genome sequencing · Comparative genomics · Population genomics · Pangenome

## Introduction

The genus *Colletotrichum*, in the phylum Ascomycota, comprises over 200 species that have been subclustered into fifteen species complexes (Talhinhas and Baroncelli 2021). Anthracnose caused by *Colletotrichum* spp. is an economically important disease on many ornamental plants as well as important grain, vegetable, and fruit crops (Cannon et al. 2012). In maize, anthracnose stalk rot caused by *C. gramini­cola* is responsible for 3.63% of estimated yield loss in the United States and Canada (Savary et al. 2019). With the total maize production in the United States and Ontario, Canada,

from 2012 to 2015 valued at over 244 billion dollars, even if only a low percentage of production is lost, the economic loss would exceed a billion dollars (Mueller et al. 2016). Anthracnose, caused by multiple *Colletotrichum* spp., is one of the most devastating global diseases of strawberry (Baroncelli et al. 2015; Freeman and Katan 1997; Jayawardena et al. 2016a). Due to intensive use of quinone-­outside inhibitor (QoI) fungicides, QoI resistance frequently occurs in strawberry-infecting *Colletotrichum* populations, which makes controlling this disease much more difficult (Forcelini et al. 2016; Inada et al. 2008). In Africa, coffee berry disease (CBD) caused by *C. kahawae* poses a threat to Arabica coffee production, potentially resulting in yield losses of 70–80%, in the absence of effective control measures (Silva et al. 2006). Although currently confined to African countries, there is a serious risk of dispersal to other Arabica coffee cultivation regions (Batista et al. 2017).

*Colletotrichum* spp. tend to exhibit a relatively high degree of host specificity. For example, *C. gramini­cola* is restricted to infecting gramineous monocots (Crouch and Beirn 2009). Likewise, *C. higginsianum* infects Brassicaceae, including *Arabidopsis thaliana* (O’Connell et

✉ Ayako Tsushima  
ayako.tsushima@omu.ac.jp

<sup>1</sup> Graduate School of Agriculture, Osaka Metropolitan University, Sakai, Japan

<sup>2</sup> Center for Sustainable Resource Science, RIKEN, Yokohama, Japan

<sup>3</sup> Graduate School of Science, The University of Tokyo, Bunkyo, Japan

al. 2004). In contrast, members of the *C. gloeosporioides* species complex tend to be post-harvest pathogens on a wide range of fruits, including avocado, banana, mango, coffee, and strawberry (Hyde et al. 2009). Whatever their host range, the majority of *Colletotrichum* spp. employ a hemibiotrophic lifestyle, characterized by the sequential development of a series of specialized cell types (Münch et al. 2008). The initial biotrophic phase is characterized by the development of bulbous primary hyphae in living host cells following penetration from melanized structures called appressoria. Host cell death is later induced in a subsequent necrotrophic phase, which is characterized by the production of filamentous secondary hyphae. Previous studies showed that hemibiotrophic infection in *Colletotrichum* spp. is orchestrated by various molecules: small secreted proteins called effectors manipulating the plant immune system (Irieda et al. 2019; Kleemann et al. 2012), carbohydrate-active enzymes (CAZymes) degrading host cell wall (Ben-Daniel et al. 2012; Yakoby et al. 2000), and secondary metabolites enhancing rigidity of appressoria or inhibiting plant hormone signaling (Dallery et al. 2020; Ludwig et al. 2014). Unlike obligate biotrophic pathogens (Spanu and Panstruga 2017), *Colletotrichum* spp. can be cultured axenically and are amenable to genetic manipulation, such as transformation and targeted knock-out mutagenesis (De Groot et al. 1998; Rikkerink et al. 1994; Rodriguez and Yoder 1987). Moreover, recently established CRISPR/Cas9 and marker recycling systems have made experimental designs more flexible and accessible (Kumakura et al. 2019; Nakamura et al. 2019; Yamada et al. 2021).

The rapid drop in the cost of next-generation genome sequencing has led to a rapid expansion of publicly available genomic information, with the number of GenBank accessions increasing at an annual rate of about 40% (Sayers et al. 2019). *Colletotrichum* spp. are no exception to this trend. Due to agricultural and general scientific interest, more than a hundred *Colletotrichum* spp. genome assemblies have been generated over the last decade. In this review, we examine the literature covering research that has contributed to the development and utilization of those genomic resources, and then we propose future perspective uses of genomic data.

## The dawn of the *Colletotrichum* genomics

The genomics of *Colletotrichum* spp. began with the release of *C. higginsianum* IMI349063 and *C. graminicola* M1.001 genome assemblies generated by a whole-genome shotgun (WGS) approach (O’Connell et al. 2012). Subsequently, the genome assemblies of *C. orbiculare* 104-T, *C. fructicola* Nara-gc5 (considered as *C. gloeosporioides*

when the genome was sequenced), and *C. gloeosporioides* Cg-14 were also published (Alkan et al. 2013; Gan et al. 2013). Comparative genomics studies using these genomic resources revealed pathogenicity-related gene repertoires, including effector candidate genes, CAZyme encoding genes, and secondary metabolite gene clusters. Furthermore, comprehensive gene annotations made RNA-seq analyses for understanding transcriptome dynamics possible. Those studies demonstrated the sequential transcriptome changes that occur during the transition from one infection stage to another (Gan et al. 2013; O’Connell et al. 2012), as well as the regulation of gene expression by the pH-responsive transcription factor *pacC* (Alkan et al. 2013). The increasing number of publicly accessible host plant genome assemblies enabled performing dual RNA-seq, in which transcriptomic changes in both host and pathogen are simultaneously analyzed. Using this approach, Alkan et al. revealed concurrent tomato and *C. gloeosporioides* alteration in gene expression during infection (Alkan et al. 2015).

## Acceleration of gene characterization

The availability of genomic information facilitated forward genetic screening techniques using *Colletotrichum* spp. *Rhizobium radiobacter* (*Agrobacterium tumefaciens*)-mediated T-DNA insertional mutagenesis has been widely used to generate *Colletotrichum* spp. mutants to identify genes involved in pathogenicity (Tsuji et al. 2003; Takahara et al. 2004). Before the availability of genomic information, researchers employed a method called “genomic walking”, where T-DNA flanking sequences are determined by thermal asymmetric interlaced polymerase chain reaction (PCR) or inverse PCR and then used to physically isolate a cosmid-type genomic clone that contains the sequences (Fujihara et al. 2010; Huser et al. 2009). In the post-genomic era, researchers can computationally search T-DNA flanking sequences against genome assemblies, which makes the identification of mutated genes relatively easy (Harata and Kubo 2014; Korn et al. 2015).

The combination of effective genetic manipulation tools and genomic information made it feasible to conduct reverse genetics experiments with *Colletotrichum* spp., which yielded new insights into infection mechanisms. Reverse genetics studies using *C. orbiculare* have identified many of the signaling components required for infection, such as CoBFA1, CoTEM1, CoWHI2, and CoPSR1 (Fukada and Kubo 2015; Harata et al. 2016), homologs of the budding yeast *Saccharomyces cerevisiae* genes that are involved in cell cycle regulation and septum formation. Interestingly, the functions of those genes differ between *C. orbiculare* and other fungi, including *S. cerevisiae*, emphasizing the

utility of the reverse genetics approach for illuminating the specific functional adaptations of genes in *Colletotrichum* spp. Genomic information from *Colletotrichum* spp. and other fungi was also used to explore the biological roles of proteins without known functional domains. Although the functions of hypothetical proteins including many putative effectors are hard to predict, comparative genomics analyses can provide valuable clues to narrow down the possible roles of genes of interest. Based on the hypothesis that *Colletotrichum* spp. use a set of common effectors during infection to support their hemibiotrophic lifestyle, Tsushima et al. explored conserved effector candidates in the genus *Colletotrichum* by comparing 24 ascomycete genomes. This study identified that a conserved effector candidate with no known functional domains, CEC3, induces nuclear expansion and host cell death (Tsushima et al. 2021). Many comparative genomics studies have provided lists of effector candidates categorized by their conservation patterns, for example, having no homolog in other *Colletotrichum* spp. (species-specific) or other genera (genus-specific) (Baroncelli et al. 2016; Bouffleur et al. 2021; Gan et al. 2016). Yet, the majority of them remain functionally uncharacterized. Experimental validation of these effector candidates is likely to be a focus of future studies.

*Colletotrichum* spp. genome assemblies also act as reference genomes for mapping next-generation sequencing (NGS) reads, they are used to assist in the identification of single nucleotide polymorphisms (SNPs), which can be used as high-density genetic markers. Bhadauria et al. generated a genome assembly of *C. lentis* and performed quantitative trait locus (QTL) mapping, then successfully identified a virulence-governing minichromosome (Bhadauria et al. 2019). QTL mapping requires crossing parental lines with different phenotypes to generate progeny, which is only possible for a limited number of *Colletotrichum* spp. where a crossing methodology has been established (Armstrong-Cho and Banniza 2006). Even if mating between different isolates is difficult under laboratory conditions, it is possible to conduct a genome-wide association study (GWAS) utilizing natural variations among unrelated individuals. A GWAS using 30 *C. kahawae* isolates identified four candidate genes that are potentially involved in signaling, detoxification, and gene expression, which may contribute to virulence on coffee berries (Vieira et al. 2019). However, for supposedly asexual *Colletotrichum* spp. (Wilson et al. 2021), GWAS settings may need to be adjusted because near-clonal genetic backgrounds can limit the power of the study due to reduced markers that are distributed by recombination (Plissonneau et al. 2017).

## Detection of adaptation signals on genomes

*Colletotrichum* spp. thrive in a variety of niches. Hence, they should have evolved distinct gene sets as a consequence of environmental adaptation. In this section, we discuss how genome data have been used to detect genes associated with survival in specific niches.

Although most copy number variations (CNVs) of genes are assumed to be detrimental, they can increase fitness by altering expression via dosage effects, or by compensating for deleterious effects of loss-of-function mutations, particularly under stressful conditions or in perturbed environments (Katju and Bergthorsson 2013). O'Connell et al. found that the dicot-infecting *C. higginsianum* encodes more CAZymes involved in pectin degradation than the monocot-adapted pathogen *C. graminicola*, suggesting that there has been adaptation to the differences in their host cell composition; eudicot cells accumulate more pectin than monocot cells (O'Connell et al. 2012). There are more genes encoding pectin and hemicellulose-degrading CAZymes in members of the *C. gloeosporioides* and *C. acutatum* species complexes compared to other members of the genus (Baroncelli et al. 2016; Gan et al. 2016). Despite belonging to phylogenetically separate branches within the genus, these two species complexes include important postharvest pathogens, suggesting a convergent gene expansion associated with their common infection strategy. Genes encoding secondary metabolite biosynthesis-related proteins are significantly enriched in plant growth-promoting *C. tofieldiae* than in pathogenic *Colletotrichum* spp., suggesting that *C. tofieldiae*-specific secondary metabolites are responsible for the beneficial endophytic interactions with host plants (Hacquard et al. 2016).

An evolutionary arms race between hosts and pathogens generates strong positive selection on both parties, a process that leaves imprints on the genes involved (Möller and Stukenbrock 2017). A classical test to detect these signatures is based on the estimated ratio of non-synonymous to synonymous mutations (dN/dS also known as  $\omega$ ) (Goldman and Yang 1994; Nei and Gojobori 1986). Using this calculation, Rech et al. explored genomic variations among eight different *C. graminicola* isolates and found that CDSs encoding secondary metabolites and putative effectors tend to have higher dN/dS ratios, indicating positive selection (Rech et al. 2014). Similarly, comparative genomics analysis using six different *Colletotrichum* spp. showed that genes encoding predicted secreted proteins, which include effector candidates, are often enriched for positively selected genes compared to other genes (Gan et al. 2016). The genes under positive selection identified in these two studies include previously characterized effector genes, *CgEPI* in *C. graminicola* and homologs of *ChELP1* in *C. higginsianum* (Vargas

et al. 2016; Takahara et al. 2016), establishing the utility of dN/dS analysis for detecting effector genes that facilitate pathogenicity.

CNV and dN/dS ratio analyses usually target gene groups or homologous pairs. However, recently emerged genes may also play significant roles in the occupation of novel niches, as has been demonstrated in effector genes, which are often restricted in certain lineages (Fouché et al. 2018). Genus- or species-specific genes have been well-documented by comparing multiple *Colletotrichum* and other fungal lineages (Buiate et al. 2017; Rao and Nandineni 2017). Moreover, recent studies identified strain-specific genes (Gan et al. 2021; Hsieh et al. 2022; Tsushima et al. 2019). New genes are thought to arise largely from two mechanisms, represented by de novo evolution from noncoding sequences, and neofunctionalization of genes obtained through genome duplication or horizontal gene transfer (HGT) (Tautz and Domazet-Lošo 2011). HGT events in *Colletotrichum* spp. have been reported for movement of the subtilisin gene from plants to the *Colletotrichum* lineage, and in accession of the cercosporin toxin biosynthesis gene cluster across the plant pathogenic fungi, including *Colletotrichum* spp. (Armijos Jaramillo et al. 2013; De Jonge et al. 2018). Although it is unclear if these horizontally transferred genes have undergone neofunctionalization once established within a *Colletotrichum* genome, up-regulation of these genes during infection suggests their contribution to disease development.

It is obvious that the quality and quantity of genomic information are crucial to conducting comparative genomics studies. Dallery et al. identified that a previous *C. higginsianum* genome assembly had 2,699 split gene models and 2,289 missing gene models, which were recovered in the latest chromosome-level genome assembly (Dallery et al. 2017). An important caveat, then, is that incomplete or disrupted chromosomal regions can diminish the utility or reliability of comparative genomics analyses. In addition, successful genome analysis relies on the accuracy of gene annotations. Most fungal genome sequencing projects employ gene annotation pipelines to generate gene models using transcriptome data and/or known homologous protein sequences as guides for prediction. However, the identification of accurate gene models using automated methods is still challenging, especially for less conserved genes, such as orphan genes whose presence is recognized in a single species (Li et al. 2022). Indeed, some studies reported transcript variants from *Colletotrichum* spp. that differ from their original models (Kumakura et al. 2021; Schliebner et al. 2014; Tsushima et al. 2021). We should notice that predicted gene models for non-model fungi, including *Colletotrichum* spp., may lack experimental support, and that different annotation methods could produce variations among gene models. Since the first few *Colletotrichum* spp.

genomics studies could only access a handful of fragmented genome assemblies (Alkan et al. 2013; Gan et al. 2013; O'Connell et al. 2012), it would be interesting to readdress the same questions with the abundant, high-quality genomic resources now available.

## Long-read sequencing and structural genomic variations

The advent of long-read sequencing technologies such as PacBio and Oxford Nanopore has revolutionized genomic studies. During de novo genome assembly using a few hundred-bp NGS reads, repetitive DNA sequences containing, for example, transposable elements (TEs) that are longer than the read length, can lead to gaps in the assembly, or misassembled rearrangements (Treangen and Salzberg 2011). This resulted in repeat sequence contents that were often underrepresented in genome assemblies (Alkan et al. 2011). Long-read sequencing technologies solved these problems by producing more than several tens of kb reads that span repeat-rich regions and generate highly contiguous genome assemblies. Using a combination of PacBio long-read sequencing and optical mapping, a chromosome-level genome assembly of *C. higginsianum* (Zampounis et al. 2016) produced some remarkable findings, including identification of the association between TEs and effector candidate genes, or secondary metabolite gene clusters, a dispensable TE-rich small chromosome required for virulence, and large-scale genomic rearrangements mediated by TEs (Dallery et al. 2017; Plaumann et al. 2018; Tsushima et al. 2019). Recently, Gan et al. also reported that large-scale genomic rearrangements and multi-copy effector candidate gene clusters are frequently associated with repeat sequences such as telomeres and TEs within the *C. gloeosporioides* species complex (Gan et al. 2021). Historically, many studies have suggested the importance of repeat sequences to generate genomic variations in plant pathogenic fungi (Chuma et al. 2003; Crouch et al. 2008; Ikeda et al. 2002), yet these findings are often restricted to specific repeat sequence families, due to the lack of the comprehensive genomic information. A comparison of chromosome-level assemblies illustrates a generalized view of the role of repeat sequences in genomic evolution at the single-nucleotide resolution. Detection of structural genomic changes has raised the question of how repeat sequences contribute to pathogen fitness. When existing genomes are compared, identification of a responsible genomic variation(s) for a given phenotype is often difficult because of the high background noise within a natural population. Further analyses using chromosome modification techniques, such as deletion or transmission of chromosomes (He et al. 1998;

Plaumann et al. 2018) should help to artificially reproduce structural genomic changes and to examine their effects on *Colletotrichum* spp. pathogenicity.

## Harnessing sequence data for species identification and diagnosis

Accumulation of genomic information has made phylogenetic analysis and species identification much more robust. A *Colletotrichum* isolate infecting shiso, or beefsteak plant (*Perilla frutescens*), was previously classified as *C. destructivum* based on morphology and the internal transcribed spacer (ITS) sequence (Kawaradani et al. 2008). However, phylogenetic analysis using concatenated multi-locus sequence data represented by ITS and four housekeeping genes from the newly-obtained genome assembly identified this isolate as a novel species, *C. shiso* (Gan et al. 2019). Because ITS sequence frequently does not provide enough resolution for species identification, multi-locus phylogenetic analysis is commonly used for taxonomic placement of *Colletotrichum* spp. as a supplement to morphological examination (Talhinhas and Baroncelli 2021). However, in general, there are no standard protein-coding genes for fungal species identification (Houbraken et al. 2021). The sequence sets used for multi-locus phylogenetic analyses using *Colletotrichum* spp. differ depending on the efficacy of each locus to resolve species delimitation in individual species complexes (Jayawardena et al. 2016b). Phylogenetic analysis using universal single-copy orthologs across genomes (Shen et al. 2020), or average nucleotide identity (ANI) analysis, which has been used extensively in bacterial taxonomic assignments (Ciufo et al. 2018), may be used with great effect to assist with *Colletotrichum* species characterization and identification.

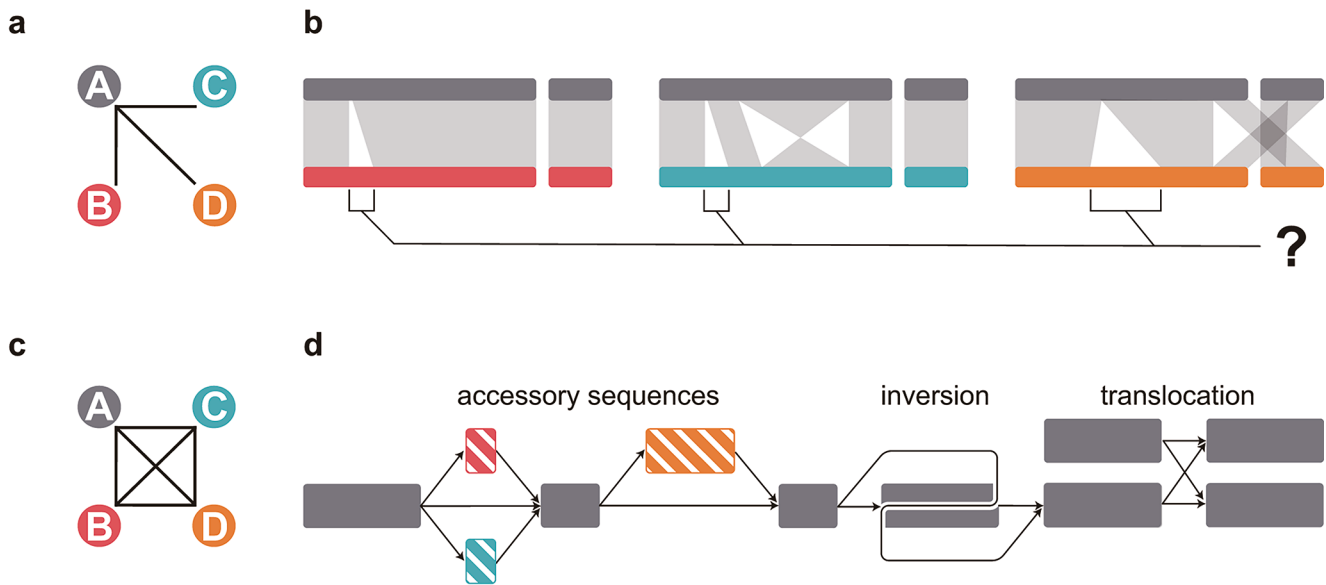
Publicly-available genomic resources of *Colletotrichum* spp. can also be used to generate diagnostic markers. Diagnosis of a causal pathogen is important to understand the economic impact caused by that pathogen and to take appropriate disease control measures. Yet, classification based on morphology and host species is less definitive, especially within a species complex or showing a similar host range. To solve this problem, Gan et al. developed diagnostic PCR makers to distinguish the four different members of the *C. gloeosporioides* species complex by comparing the genomes (Gan et al. 2017). Analysis using these PCR markers showed that *C. fructicola* has been the predominant species causing strawberry anthracnose in Chiba, Japan (Gan et al. 2017) and that *C. fructicola* was detected in 5.7% of tested weed leaves around strawberry nurseries in Nara, Japan, which could be an inoculum source of the disease (Hirayama et al. 2018). Apart from species diagnosis, the

agricultural industry has a great need for inspecting other aspects of pathogens, including lineage, aggressiveness, or fungicide sensitivity. Population genomics can aid in monitoring them all together by analyzing high-resolution genotypic data generated by mapping NGS reads to reference genomes. For example, the field pathogenomics approach, which utilizes RNA-seq data of infected host tissues from fields, provides a way to identify phylogenetic relationships between samples (Hubbard et al. 2015; Islam et al. 2016), to estimate a race based on genetic proximity to known isolates (Lewis et al. 2018; Tsushima et al. 2022), or to evaluate mutations in fungicide target genes (Cook et al. 2021). A wealth of sequence data should make the diagnosis of *Colletotrichum* spp. more accurate, flexible, and tailored in the future.

## Toward *Colletotrichum* pangenomics

Genome assemblies had at one time been generated for a representative isolate of each species. However, it is now easier to obtain and compare several genome assemblies from a single species due to the low cost of sequencing. This technological advancement, combined with computational power, has generated the concept of pangenome, which characterizes the entire set of genomic sequences within a phylogenetic clade of interest (e.g. species) (Vernikos et al. 2015) (Fig. 1). Pangenomic studies determine the genomic diversity among all available datasets, spanning from highly conserved core sequences to sporadically arisen accessory sequences. Identification of such rapidly evolving accessory sequences is particularly important for the study of plant pathogens because they could contribute to selective advantages in the arms race with host plants (Badet and Croll 2020). Although reference-based genome alignments are practical, scalable, and demand less computational power, they can only detect genomic variations that are present in a particular reference genome that may be chosen arbitrarily (Eizenga et al. 2020). This is an issue for analyzing fungal phytopathogen genomes because previous studies show that these genomes are highly plastic at the chromosomal level (Faino et al. 2016; Li et al. 2019; Tsushima et al. 2019). In the near future, comparing many genome assemblies will be possible to investigate specific selective pressures against structural variations.

Pangenome construction requires huge genomic datasets. How can these massive datasets be produced, archived, and examined? The most straightforward way is to generate genome assemblies by individual research groups. However, there is a limit to how much a single team can handle. A more realistic solution would be to harness open data resources with a common architecture and coordinating



**Fig. 1** The concept of pangenome. **a** In reference-based genome alignments, one genome (A) is chosen as a reference, then the others (B–D) are compared with it. **b** One-to-one alignments cannot represent mutual relationships of unaligned genomic regions. Enough similar sequences are linked between the reference (top) and another (bottom) in each diagram. **c** Pangenomics reciprocally compares all

authority. Sequence data deposited in public databases provide a great opportunity to investigate the genomic atlas of fungi, including *Colletotrichum* spp. Despite the benefits of analyzing public data, as users, we should be aware that those data were obtained in different ways, and that filtering inappropriate datasets is often required to reduce project-dependent biases (Sielemann et al. 2020). As sequence depositors, providing detailed, accurate, and timeless metadata along with the sequence is essential to assist integrative analyses by other researchers, and to eventually maximize the value of the data (Wilkinson et al. 2016). A possible bottleneck for expanding the available *Colletotrichum* spp. genomic data would be to find *Colletotrichum* strains that asymptotically infect plants, or that grow in organic matter without hosts (Silva et al. 2017). We can easily overlook non-pathogenic *Colletotrichum* strains, although they may harbor intriguing traits, like promotion of plant growth (Hiruma et al. 2016; Ye et al. 2020). The whole-metagenome shotgun sequencing approach holds promises for obtaining novel genomic information of *Colletotrichum* spp. in nature. Assembling whole-metagenome sequence reads is still challenging due to low sequencing depth, especially for eukaryotes (Bandla et al. 2020; Regalado et al. 2020). However, some studies successfully recovered fungal metagenome-assembled genomes (MAGs) (West et al. 2018; Peng et al. 2021). In the future, application of long-read sequencing technologies could improve the quality and quantity of eukaryotic MAGs. To capture the global image

of the genomes. **d** Many-to-many sequence alignments clarify the entire sequence repertoire without a reference bias. A sequence graph illustrates core sequences (solid rectangles) and accessory sequences (stripe rectangles) as well as structural variations (inversion and translocation)

of genomic diversity of *Colletotrichum* spp., we advocate for collecting and analyzing a multitude of genomes, even within a single species.

**Acknowledgements** This work was supported in part by KAKENHI (22H00364 to KS).

## Declarations

**Conflict of interest** The authors have no conflicts of interest to declare.

**Human and animal rights** This article does not contain any studies with human participants or animals performed by any of the authors.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Alkan C, Sajjadian S, Eichler EE (2011) Limitations of next-generation genome sequence assembly. *Nat Methods* 8:61–65
- Alkan N, Meng X, Friedlander G et al (2013) Global Aspects of *pacC* Regulation of Pathogenicity Genes in *Colletotrichum gloeosporioides* as Revealed by Transcriptome Analysis. *Mol Plant-Microbe Interact* 26:1345–1358
- Alkan N, Friedlander G, Ment D et al (2015) Simultaneous transcriptome analysis of *Colletotrichum gloeosporioides* and tomato fruit pathosystem reveals novel fungal pathogenicity and fruit defense strategies. *New Phytol* 205:801–815
- Armijos Jaramillo VD, Vargas WA, Sukno SA, Thon MR (2013) Horizontal Transfer of a Subtilisin Gene from Plants into an Ancestor of the Plant Pathogenic Fungal Genus *Colletotrichum*. *PLoS ONE* 8:e59078
- Armstrong-Cho CL, Banniza S (2006) *Glomerella truncata* sp. nov., the teleomorph of *Colletotrichum truncatum*. *Mycol Res* 110:951–956
- Badet T, Croll D (2020) The rise and fall of genes: origins and functions of plant pathogen pangenomes. *Curr Opin Plant Biol* 56:65–73
- Bandla A, Pavagadhi S, Sridhar Sudarshan A et al (2020) 910 metagenome-assembled genomes from the phytobiomes of three urban-farmed leafy Asian greens. *Sci Data* 7:278
- Baroncelli R, Zapparata A, Sarrocco S et al (2015) Molecular diversity of Anthracnose Pathogen populations associated with UK Strawberry production suggests multiple introductions of three different *Colletotrichum* species. *PLoS ONE* 10:e0129140
- Baroncelli R, Amby DB, Zapparata A et al (2016) Gene family expansions and contractions are associated with host range in plant pathogens of the genus *Colletotrichum*. *BMC Genomics* 17:555
- Batista D, Silva DN, Vieira A et al (2017) Legitimacy and implications of reducing *Colletotrichum kahawae* to subspecies in plant pathology. *Front Plant Sci* 7:2051
- Ben-Daniel BH, Bar-Zvi D, Tsror Lahkim L (2012) Pectate lyase affects pathogenicity in natural isolates of *Colletotrichum* coccodes and in *pelA* gene-disrupted and gene-overexpressing mutant lines. *Mol Plant Pathol* 13:187–197
- Bhadauria V, Maclachlan R, Pozniak C et al (2019) Genetic map-guided genome assembly reveals a virulence-governing minichromosome in the lentil anthracnose pathogen *Colletotrichum lentis*. *New Phytol* 221:431–445
- Bouffleur TR, Massola Júnior NS, Tikami Í et al (2021) Identification and comparison of *Colletotrichum* secreted effector candidates reveal two independent lineages pathogenic to Soybean. *Pathogens* 10:1520
- Buiate EAS, Xavier KV, Moore N et al (2017) A comparative genomic analysis of putative pathogenicity genes in the host-specific sibling species *Colletotrichum graminicola* and *Colletotrichum sublineola*. *BMC Genomics* 18:67
- Cannon PF, Damm U, Johnston PR, Weir BS (2012) *Colletotrichum* – current status and future directions. *Stud Mycol* 73:181–213
- Chuma I, Tosa Y, Taga M et al (2003) Meiotic behavior of a super-numerary chromosome in *Magnaporthe oryzae*. *Curr Genet* 43:191–198
- Ciufo S, Kannan S, Sharma S et al (2018) Using average nucleotide identity to improve taxonomic assignments in prokaryotic genomes at the NCBI. *Int J Syst Evol Microbiol* 68:2386–2392
- Cook NM, Chng S, Woodman TL et al (2021) High frequency of fungicide resistance-associated mutations in the wheat yellow rust pathogen *Puccinia striiformis* f. sp. *tritici*. *Pest Manag Sci* 77:3358–3371
- Crouch JA, Glasheen BM, Giunta MA et al (2008) The evolution of transposon repeat-induced point mutation in the genome of *Colletotrichum cereale*: Reconciling sex, recombination and homoplasmy in an “asexual” pathogen. *Fungal Genet Biol* 45:190–206
- Crouch JA, Beirn LA (2009) Anthracnose of cereals and grasses. *Fungal Divers* 39:19–44
- Dallery JF, Lapalu N, Zampounis A et al (2017) Gapless genome assembly of *Colletotrichum higginsianum* reveals chromosome structure and association of transposable elements with secondary metabolite gene clusters. *BMC Genomics* 18:667
- Dallery JF, Zimmer M, Halder V et al (2020) Inhibition of jasmonate-mediated plant defences by the fungal metabolite higginsianin B. *J Exp Bot* 71:2910–2921
- De Groot MJA, Bundock P, Hooykaas PJJ et al (1998) *Agrobacterium tumefaciens*-mediated transformation of filamentous fungi. *Nat Biotechnol* 16:839–842
- De Jonge R, Ebert MK, Huijt-Roehl CR et al (2018) Gene cluster conservation provides insight into cercosporin biosynthesis and extends production to the genus *Colletotrichum*. *Proc Natl Acad Sci U S A* 115:E5459–E5466
- De Silva DD, Crous PW, Ades PK et al (2017) Life styles of *Colletotrichum* species and implications for plant biosecurity. *Fungal Biol Rev* 31:155–168
- Eizenga JM, Novak AM, Sibbesen JA et al (2020) Pangenome graphs. *Annu Rev Genomics Hum Genet* 21:139–162
- Faino L, Seidl MF, Shi-Kunne X et al (2016) Transposons passively and actively contribute to evolution of the two-speed genome of a fungal pathogen. *Genome Res* 26:1091–1100
- Forcelini BB, Seijo TE, Amiri A, Peres NA (2016) Resistance in strawberry isolates of *Colletotrichum acutatum* from Florida to quinone-outside inhibitor fungicides. *Plant Dis* 100:2050–2056
- Fouché S, Plissonneau C, Croll D (2018) The birth and death of effectors in rapidly evolving filamentous pathogen genomes. *Curr Opin Microbiol* 46:34–42
- Freeman S, Katan T (1997) Identification of *Colletotrichum* species responsible for anthracnose and root necrosis of strawberry in Israel. *Phytopathology* 87:516–521
- Fujihara N, Sakaguchi A, Tanaka S et al (2010) Peroxisome biogenesis factor PEX13 is required for appressorium-mediated plant infection by the anthracnose fungus *Colletotrichum orbiculare*. *Mol Plant-Microbe Interact* 23:436–445
- Fukada F, Kubo Y (2015) *Colletotrichum orbiculare* regulates cell cycle G1/S progression via a two-component GAP and a GTPase to establish plant infection. *Plant Cell* 27:2530–2544
- Gan P, Ikeda K, Irieda H et al (2013) Comparative genomic and transcriptomic analyses reveal the hemibiotrophic stage shift of *Colletotrichum* fungi. *New Phytol* 197:1236–1249
- Gan P, Narusaka M, Kumakura N et al (2016) Genus-wide comparative genome analyses of *Colletotrichum* species reveal specific gene family losses and gains during adaptation to specific infection lifestyles. *Genome Biol Evol* 8:1467–1481
- Gan P, Nakata N, Suzuki T, Shirasu K (2017) Markers to differentiate species of anthracnose fungi identify *Colletotrichum fructicola* as the predominant virulent species in strawberry plants in Chiba Prefecture of Japan. *J Gen Plant Pathol* 83:14–22
- Gan P, Tsushima A, Hiroshima R et al (2019) *Colletotrichum shioi* sp. nov., an anthracnose pathogen of *Perilla frutescens* in Japan: molecular phylogenetic, morphological and genomic evidence. *Sci Rep* 9:13349
- Gan P, Hiroshima R, Tsushima A et al (2021) Telomeres and a repeat-rich chromosome encode effector gene clusters in plant pathogenic *Colletotrichum* fungi. *Environ Microbiol* 23:6004–6018
- Goldman N, Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol* 11:725–736

- Hacquard S, Kracher B, Hiruma K et al (2016) Survival trade-offs in plant roots during colonization by closely related beneficial and pathogenic fungi. *Nat Commun* 7:11362
- Harata K, Kubo Y (2014) Ras GTPase activating protein coir1 is involved in infection-related morphogenesis by regulating cAMP and MAPK signaling pathways through CoRas2 in *Colletotrichum orbiculare*. *PLoS ONE* 9:e109045
- Harata K, Nishiuchi T, Kubo Y (2016) *Colletotrichum orbiculare* *WHI2*, a yeast stress-response regulator homolog, controls the biotrophic stage of hemibiotrophic infection through TOR signaling. *Mol Plant-Microbe Interact* 29:468–483
- He C, Rusu AG, Poplawski AM et al (1998) Transfer of a supernumerary chromosome between vegetatively incompatible biotypes of the fungus *Colletotrichum gloeosporioides*. *Genetics* 150:1459–1466
- Hirayama Y, Asano S, Okayama K et al (2018) Weeds as the potential inoculum source of *Colletotrichum fructicola* responsible for strawberry anthracnose in Nara, Japan. *J Gen Plant Pathol* 84:12–19
- Hiruma K, Gerlach N, Sacristán S et al (2016) Root endophyte *Colletotrichum tofieldiae* confers plant fitness benefits that are phosphate status dependent. *Cell* 165:464–474
- Houbraken J, Visagie CM, Frisvad JC (2021) Recommendations to prevent taxonomic misidentification of genome-sequenced fungal strains. *Microbiol Resour Announc* 10:e0107420
- Hsieh DK, Chuang SC, Chen CY et al (2022) Comparative genomics of three *Colletotrichum scovillei* strains and genetic analysis revealed genes involved in fungal growth and virulence on Chili Pepper. *Front Microbiol* 13:818291
- Hubbard A, Lewis CM, Yoshida K et al (2015) Field pathogenomics reveals the emergence of a diverse wheat yellow rust population. *Genome Biol* 16:23
- Huser A, Takahara H, Schmalenbach W, O'Connell R (2009) Discovery of pathogenicity genes in the crucifer anthracnose fungus *Colletotrichum higginsianum*, using random insertional mutagenesis. *Mol Plant Microbe Interact* 22:143–156
- Hyde KD, Cai L, Cannon PF et al (2009) *Colletotrichum*-names in current use. *Fungal Divers* 39:147–182
- Ikeda KI, Nakayashiki H, Kataoka T et al (2002) Repeat-induced point mutation (RIP) in *Magnaporthe grisea*: implications for its sexual cycle in the natural field context. *Mol Microbiol* 45:1355–1364
- Inada M, Ishii H, Chung WH et al (2008) Occurrence of strobilurin-resistant strains of *Colletotrichum gloeosporioides* (*Glomerella cingulata*), the causal fungus of strawberry anthracnose (in Japanese with English summary). *Jpn J Phytopathol* 74:114–117
- Irieda H, Inoue Y, Mori M et al (2019) Conserved fungal effector suppresses PAMP-triggered immunity by targeting plant immune kinases. *Proc Natl Acad Sci U S A* 116:496–505
- Islam MT, Croll D, Gladieux P et al (2016) Emergence of wheat blast in Bangladesh was caused by a South American lineage of *Magnaporthe oryzae*. *BMC Biol* 14:84
- Jayawardena RS, Huang JK, Jin BC et al (2016a) An account of *Colletotrichum* species associated with strawberry anthracnose in China based on morphology and molecular data. *Mycosphere* 7:1147–1163
- Jayawardena RS, Hyde KD, Damm U et al (2016b) Notes on currently accepted species of *Colletotrichum*. *Mycosphere* 7:1192–1260
- Katju V, Bergthorsson U (2013) Copy-number changes in evolution: Rates, fitness effects and adaptive significance. *Front Genet* 4:273
- Kawaradani M, Nishimura A, Moriwaki J et al (2008) Anthracnose of perilla (*Perilla ocymoides* L.) caused by *Colletotrichum destructivum* (in Japanese with English summary). *Jpn J Phytopathol* 74:335–339
- Kleemann J, Rincon-Rivera LJ, Takahara H et al (2012) Sequential delivery of host-induced virulence effectors by Appressoria and Intracellular Hyphae of the phytopathogen *Colletotrichum higginsianum*. *PLoS Pathog* 8:e1002643
- Korn M, Schmidpeter J, Dahl M et al (2015) A genetic screen for pathogenicity genes in the hemibiotrophic fungus *Colletotrichum higginsianum* identifies the plasma membrane proton pump Pma2 required for host penetration. *PLoS ONE* 10:e0125960
- Kumakura N, Ueno A, Shirasu K (2019) Establishment of a selection marker recycling system for sequential transformation of the plant-pathogenic fungus *Colletotrichum orbiculare*. *Mol Plant Pathol* 20:447–459
- Kumakura N, Singkaravanit-Ogawa S, Gan P et al (2021) Guanosine-specific single-stranded ribonuclease effectors of a phytopathogenic fungus potentiate host immune responses. *bioRxiv* 2021.10.13.464185. <https://doi.org/10.1101/2021.10.13.464185>
- Lewis CM, Persoons A, Bebbler DP et al (2018) Potential for re-emergence of wheat stem rust in the United Kingdom. *Commun Biol* 1:13
- Li F, Upadhyaya NM, Sperschneider J et al (2019) Emergence of the Ug99 lineage of the wheat stem rust pathogen through somatic hybridisation. *Nat Commun* 10:5068
- Li J, Singh U, Bhandary P et al (2022) Foster thy young: enhanced prediction of orphan genes in assembled genomes. *Nucleic Acids Res* 50:e37
- Ludwig N, Löhner M, Hempel M et al (2014) Melanin is not required for Turgor generation but enhances cell-wall rigidity in Appressoria of the corn pathogen *Colletotrichum graminicola*. *Mol Plant-Microbe Interact* 27:315–327
- Möller M, Stukenbrock EH (2017) Evolution and genome architecture in fungal plant pathogens. *Nat Rev Microbiol* 15:756–771
- Mueller DS, Wise KA, Sisson AJ et al (2016) Corn yield loss estimates due to diseases in the United States and Ontario, Canada from 2012 to 2015. *Plant Health Prog* 17:211–222
- Münch S, Lingner U, Floss DS et al (2008) The hemibiotrophic lifestyle of *Colletotrichum* species. *J Plant Physiol* 165:41–51
- Nakamura M, Okamura Y, Iwai H (2019) Plasmid-based and -free methods using CRISPR/Cas9 system for replacement of targeted genes in *Colletotrichum sansevieriae*. *Sci Rep* 9:18947
- Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3:418–426
- O'Connell R, Herbert C, Sreenivasaprasad S et al (2004) A novel *Arabidopsis-Colletotrichum* pathosystem for the molecular dissection of plant-fungal interactions. *Mol Plant-Microbe Interact* 17:272–282
- O'Connell RJ, Thon MR, Hacquard S et al (2012) Lifestyle transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses. *Nat Genet* 44:1060–1065
- Peng X, Wilken SE, Lankiewicz TS et al (2021) Genomic and functional analyses of fungal and bacterial consortia that enable lignocellulose breakdown in goat gut microbiomes. *Nat Microbiol* 6:499–511
- Plaumann P-L, Schmidpeter J, Dahl M et al (2018) A dispensable chromosome is required for virulence in the hemibiotrophic plant pathogen *Colletotrichum higginsianum*. *Front Microbiol* 9:1005
- Plissonneau C, Benevenuto J, Mohd-Assaad N et al (2017) Using population and comparative genomics to understand the genetic basis of effector-driven fungal pathogen evolution. *Front Plant Sci* 8:119
- Rao S, Nandineni MR (2017) Genome sequencing and comparative genomics reveal a repertoire of putative pathogenicity genes in chilli anthracnose fungus *Colletotrichum truncatum*. *PLoS ONE* 12:e0183567
- Rech GE, Sanz-Martín JM, Anisimova M et al (2014) Natural selection on coding and noncoding DNA sequences is associated with virulence genes in a plant pathogenic fungus. *Genome Biol Evol* 6:2368–2379



- Regalado J, Lundberg DS, Deusch O et al (2020) Combining whole-genome shotgun sequencing and rRNA gene amplicon analyses to improve detection of microbe–microbe interaction networks in plant leaves. *ISME J* 14:2116–2130
- Rikkerink EHA, Solon SL, Crowhurst RN, Templeton MD (1994) Integration of vectors by homologous recombination in the plant pathogen *Glomerella cingulata*. *Curr Genet* 25:202–208
- Rodriguez RJ, Yoder OC (1987) Selectable genes for transformation of the fungal plant pathogen *Glomerella cingulata* f. sp. *phaseoli* (*Colletotrichum lindemuthianum*). *Gene* 54:73–81
- Savary S, Willocquet L, Pethybridge SJ et al (2019) The global burden of pathogens and pests on major food crops. *Nat Ecol Evol* 3:430–439
- Sayers EW, Cavanaugh M, Clark K et al (2019) GenBank. *Nucleic Acids Res* 47:D94–D99
- Schliebner I, Becher R, Hempel M et al (2014) New gene models and alternative splicing in the maize pathogen *Colletotrichum graminicola* revealed by RNA-Seq analysis. *BMC Genomics* 15:842
- Shen XX, Steenwyk JL, LaBella AL et al (2020) Genome-scale phylogeny and contrasting modes of genome evolution in the fungal phylum Ascomycota. *Sci Adv* 6:eabd0079
- Sielemann K, Hafner A, Pucker B (2020) The reuse of public datasets in the life sciences: potential risks and rewards. *PeerJ* 8:e9954
- Silva MDC, Várzea V, Guerra-Guimarães L et al (2006) Coffee resistance to the main diseases: leaf rust and coffee berry disease. *Brazilian J Plant Physiol* 18:119–147
- Spanu PD, Panstruga R (2017) Editorial: Biotrophic plant-microbe interactions. *Front Plant Sci* 8:192
- Takahara H, Tsuji G, Kubo Y, Yamamoto M et al (2004) *Agrobacterium tumefaciens*-mediated transformation as a tool for random mutagenesis of *Colletotrichum trifolii*. *J Gen Plant Pathol* 70:93–96
- Takahara H, Hacquard S, Kombrink A et al (2016) *Colletotrichum higginsianum* extracellular LysM proteins play dual roles in appressorial function and suppression of chitin-triggered plant immunity. *New Phytol* 211:1323–1337
- Talhinhas P, Baroncelli R (2021) *Colletotrichum* species and complexes: geographic distribution, host range and conservation status. *Fungal Divers* 110:109–198
- Tautz D, Domazet-Lošo T (2011) The evolutionary origin of orphan genes. *Nat Rev Genet* 12:692–702
- Treangen TJ, Salzberg SL (2011) Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genet* 13:36–46
- Tsuji G, Fujii S, Fujihara N et al (2003) *Agrobacterium tumefaciens*-mediated transformation for random insertional mutagenesis in *Colletotrichum lagenarium*. *J Gen Plant Pathol* 69:230–239
- Tsushima A, Gan P, Kumakura N et al (2019) Genomic plasticity mediated by transposable elements in the plant pathogenic fungus *Colletotrichum higginsianum*. *Genome Biol Evol* 11:1487–1500
- Tsushima A, Narusaka M, Gan P et al (2021) The conserved *Colletotrichum* spp. effector candidate CEC3 induces nuclear expansion and cell death in plants. *Front Microbiol* 12:682155
- Tsushima A, Lewis CM, Flath K et al (2022) Wheat stem rust recorded for the first time in decades in Ireland. *Plant Pathol* 71:890–900
- Vargas WA, Sanz-Martin JM, Rech GE et al (2016) A fungal effector with host nuclear localization and dna-binding properties is required for maize anthracnose development. *Mol Plant-Microbe Interact* 29:83–95
- Vernikos G, Medini D, Riley DR, Tettelin H (2015) Ten years of pan-genome analyses. *Curr Opin Microbiol* 23:148–154
- Vieira A, Silva DN, Várzea V et al (2019) Genome-wide signatures of selection in *Colletotrichum kahawae* reveal candidate genes potentially involved in pathogenicity and aggressiveness. *Front Microbiol* 10:1374
- West PT, Probst AJ, Grigoriev IV et al (2018) Genome-reconstruction for eukaryotes from complex natural microbial communities. *Genome Res* 28:569–580
- Wilkinson MD, Dumontier M, Aalbersberg IJ et al (2016) The FAIR guiding principles for scientific data management and stewardship. *Sci Data* 3:160018
- Wilson AM, Lelwala RV, Taylor PWJ et al (2021) Unique patterns of mating pheromone presence and absence could result in the ambiguous sexual behaviors of *Colletotrichum* species. *G3* 11:jkab187
- Yakoby N, Freeman S, Dinour A et al (2000) Expression of Pectate Lyase from *Colletotrichum gloeosporioides* in *C. magna* promotes pathogenicity. *Mol Plant-Microbe Interact* 13:887–891
- Yamada K, Yamamoto T, Uwasa K et al (2021) The establishment of multiple knockout mutants of *Colletotrichum orbiculare* by CRISPR/Cas9 and Cre/loxP systems. *bioRxiv* 2021.10.24.465644. DOI: <https://doi.org/10.1101/2021.10.24.465644>
- Ye B, Wu Y, Zhai X et al (2020) Beneficial effects of endophytic fungi from the *Anoectochilus* and *Ludisia* species on the growth and secondary metabolism of *Anoectochilus roxburghii*. *ACS Omega* 5:3487–3497
- Zampounis A, Pigné S, Dallery J et al (2016) Genome sequence and annotation of *Colletotrichum higginsianum*, a causal agent of crucifer anthracnose disease. *Genome Announc* 4:e00821–e00816

**Publisher's Note** Springer Nature and the Phytopathological Society of Japan remains neutral with regard to jurisdictional claims in geographical names, published maps and institutional affiliations.