Supplementary Material

Supplementary Methods:

1. Data processing:

The main interface of this pipeline relies on a dataset object, which can be built from NCBI datasets downloaded using the NCBI toolkit. Users also have the flexibility to incorporate their own genome assemblies, provided that all necessary information is included and properly formatted. We have also provided custom scripts to assemble and update all genomes and related annotation information into a single centralized dataset object. Once built, all relevant information, including annotation data from GFF files, is stored in a local SQLite database. This enables rapid access as a dataset object without the need to repeatedly process the raw dataset.

2. Multiple alignment:

Upon receiving input, the program will query three fields: species name, genus name and assembly accession, to retrieve the corresponding accession number of the genome assembly. Subsequently, it will automatically select a high-quality annotated genome as a reference from the included species, unless the user specifies otherwise. The pipeline will initialize a BLAST[1] search against the database using all gene-coding sequences extracted from the reference genome. The results then undergo parsing and filtration to identify homologous sequences. In cases where multiple matches are identified within a single assembly, only the match with the lowest E-value will be retained. Subsequently, all homologous sequence groups that meet species-specific requirements and pass the quality control criteria will be aligned using MAFFT[2]. Regions with low sequence conservation across included assemblies will also be filtered and masked.

3. Primer design considerations for included and excluded genomes:

Initially, the program attempts to design primers for all genes. However, genes must be present in all included reference genomes to be considered for primer design. For instance, in Fig.1A, Gene d is excluded from primer design due to this reason. In cases where a gene is present in both included and excluded genomes, the aligned segment within the multiple sequence alignment from the excluded reference genome is extracted and utilized as a mispriming library into Primer3[3]. By integrating this mispriming library, the primer design aims to circumvent regions prone to mispriming, thus reducing the potential for false positives originating from excluded genomes.

4. Primer quality:

Following primer design, the quality assessment of each primer sets is conducted. This evaluation employs independent short sequence BLAST analysis to ensure that the designed primers meet the requirements for specificity and sensitivity. Sequences of the primer set are compared against all genome assemblies, with each alignment assigned a penalty score based on the amount and location of mismatches (if there is any). The program will then identify all possible amplicons and pairs of binding sites for primer sets. If both the left and right primer (and the oligo probe) have a penalty below a default threshold (stricter for assemblies that are labeled with "inclusion"), the primer set will be considered as being able to produce amplicons in these genome assemblies. Furthermore, primer sets will be filtered out if an amplicon can be detected in any of the excluded genomes.

5. Output:

In accordance with user preferences, the pipeline can generate detailed descriptions of homologous sequences and primer sets in rich text format. Additionally, a condensed table containing basic information about each primer is provided. Intermediate files—input and output for BLAST and MAFFT—are also preserved by default.

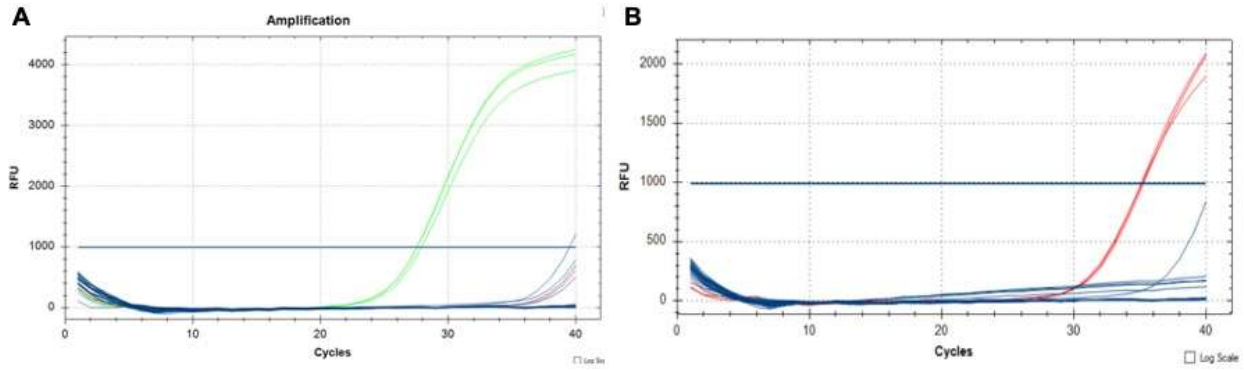
6. Computational efficiency:

The test was conducted on a single computational node with 12 cores. Taking the specific instance of *C. gattii* primer design as an example, by encompassing 23 inclusion assemblies and 1816 distinct exclusion species, 5243 groups of homologous groups were retained for alignment and primer design. For the design with Taqman internal oligo probe[4], the computational runtime spanned approximately 13 hours, yielding two sets of primers that passed the final test. For the design excluding internal oligo nucleotides, the run time was comparable, also

approximately 13 hours, resulting in the identification of 23 primers sets.

Supplementary Table 1: other control species used for Figure 1B primer testing

Number of Control species	Control pathogen species name	Number of replicates
1	Aspergillus niger	3
2	Aspergillus fumigatus	3
3	Candida albicans	3
4	Acinetobacter baumannii	3
5	Klebsiella pneumoniae	3
6	Neisseria meningitidis	3
7	Legionella pneumophila	3
8	Streptococcus pneumoniae	3
9	Haemophilus influenzae	3



Supplementary Figure 1: additional PCR primer design validation.
(A) targeted amplified genome is *C. neoformans* (green). (B) targeted amplified genome is *C. gattii*. A total of nine other infectious pathogen were used as control (See Sup Table 1). RFU: Relative Fluorescence Units.

References:

1. Altschul S F, Gish W, Miller W, Myers E W, Lipman D J. Basic Local Alignment Search Tool. Journal of Molecular Biology, 1990, 215(3): 403-410
2. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research, 2002, 30(14): 3059-3066
3. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth B C, Remm M, Rozen S G. Primer3-new capabilities and interfaces. Nucleic Acids Research, 2012, 40(15): e115
4. Navarro E, Serrano-Heras G, Castano M J, Solera J. Real-time PCR detection chemistry. Clinica Chimica Acta, 2015, 439: 231-250