Project title:	Rapid genomics approaches for downy mildews on high-value crops
Project number:	CP 195
Project leader:	Dr Thomas Wood, NIAB
Report:	Final report, March, 2020
Previous report:	N/A
Key staff:	Dr Thomas Wood Dr Beatrice Corsi Dr Anne Webb
Location of project:	NIAB
Industry Representative:	N/A
Date project commenced:	14th October 2019

DISCLAIMER

While the Agriculture and Horticulture Development Board seeks to ensure that the information contained within this document is accurate at the time of printing, no warranty is given in respect thereof and, to the maximum extent permitted by law the Agriculture and Horticulture Development Board accepts no liability for loss, damage or injury howsoever caused (including that caused by negligence) or suffered directly or indirectly in relation to information and opinions contained in or omitted from this document.

© Agriculture and Horticulture Development Board 2020. No part of this publication may be reproduced in any material form (including by photocopy or storage in any medium by electronic mean) or any copy or adaptation stored, published or distributed (by physical, electronic or other means) without prior permission in writing of the Agriculture and Horticulture Development Board, other than by reproduction in an unmodified form for the sole purpose of use as an information resource when the Agriculture and Horticulture Development Board or AHDB Horticulture is clearly acknowledged as the source, or in accordance with the provisions of the Copyright, Designs and Patents Act 1988. All rights reserved.

All other trademarks, logos and brand names contained in this publication are the trademarks of their respective holders. No rights are granted without the prior written permission of the relevant owners.

The results and conclusions in this report are based on an investigation conducted over a one-year period. The conditions under which the experiments were carried out and the results have been reported in detail and with accuracy. However, because of the biological nature of the work it must be borne in mind that different circumstances and conditions could produce different results. Therefore, care must be taken with interpretation of the results, especially if they are used as the basis for commercial product recommendations.

AUTHENTICATION

We declare that this work was done under our supervision according to the procedures described herein and that the report represents a true and accurate record of the results obtained.

Dr Thomas Wood

Programme Leader, Plant Pathology

NIAB

Signature T. Wood Date 17th March 2020

CONTENTS

GROWER SUMMARY

Headlines

- New genomics resources have been generated for basil and column stock downy mildews. These can be used in fundamental and applied studies to improve knowledge and understanding to help manage these important pathogens.
- Additional genomes for downy mildew species affecting spinach and sugar beet will be completed imminently, providing new public resources for investigating these serious pathogens.
- The downy mildew genomes have been exploited to design new (or to refine existing) isothermal diagnostics (LAMP) assays that will enhance practical capability to reduce the impact of downy mildew through rapid detection.

Background

The aim of the project was to generate *de novo* hybrid genome assemblies based on Illumina short-read and Oxford Nanopore Technology long-read chemistries to provide new, novel genomics resources for *P. destructor* on chives*, P. lamii /* additional species on mint/sage*, P. belbahrii* on sweet basil and *P. petroselini* on parsley. Subsequently we proposed to conduct a preliminary phylogenetic investigation to establish relatedness between the four downy mildew (DM) species, design new isothermal amplification (LAMP) assays to provide diagnostic support for growers, and finally to ensure the new sequence data was made available through public repositories (Genbank).

The work program was organised to address the following objectives:

i: Obtain pure cultures of target DMs and extract DNA for sequencing

ii: Produce *de novo* draft genome assemblies of four target DM species

iii: Conduct comparative genomics and phylogenetic analyses to compare gene content, genome structure/organization

iv: Design three new isothermal assays for detecting target species

v: Sequence data and assay sequences deposited in public repository so the information can be accessed and utilized effectively

Due to the timing of the project occurring after the growing season, viable samples could only be obtained for *P. belbahrii* and not the three remaining target species. After discussion with the funders, three additional downy mildews, *Peronospora effusa* affecting spinach, *Peronospora farinosa* f. sp *betae* on sugarbeet and *Peronospora matthiolae* on the ornamental species column stock (*Matthiola incana)* were selected as alternatives for the new generation of genomics resources. The basis for the selection was that *P. effusa* and *P. farinosa* f. sp *betae* are closely related and affect economically important spinach and sugarbeet / table beet, respectively; sequencing them provided an opportunity to compare for differences that may confer host-specificity or specialisation, and differences in pathogenicity for *P. effusa* where other race-types have recently been sequenced[1]. *P. matthiolae* affects high-value, clonally propagated ornamental species, and is thought to be closely related to other economically important pathogenic downy mildews on Brassica species and model plant *Arabidopsis thaliana*.

Summary

Downy mildews (DMs) are obligate oomycete pathogens that cause serious economic losses in high value crops and ornamental species such as sweet basil, chives and parsley. Control of downy mildew disease is often confounded by a lack of genetic resistance, need to minimise residues in food, continued reduction in effective fungicides and issues with effective control due to fungicide resistance. Despite their economic importance, and an established knowledge base and accompanying genomics resources for a number of major and model oomycete plants pathogens (i.e. *P. infestans, P. ultimum* and *H. arabidopsidis),* there is still a paucity of relevant tools for many major crop downy mildew pathogens. In an effort to help address this, the project has utilised next-generation sequencing technologies to rapidly and accurately create novel genomics resources for key downy mildew pathogens affecting high value horticultural, arable and ornamental crops.

We have generated *de novo* genome assemblies for *P. belbahrii* (48 Mb) and *P. matthiolae* (112 Mb) which have begun to facilitate fundamental comparative genomics studies with other downy mildews, whilst also enabling us to design new LAMP diagnostics for specific detection of column stock downy mildew to aid growers. The new assay can effectively detect mildew spores or symptomatic *M. incana* leaf tissue.

A preliminary phylogenetic analysis was conducted by comparing newly generated (ITS Internal Transcribed Spacer region) sequences from *P. matthiolae* (column stock DM) with those from *P. belbahrii* (basil DM) and publicly available sequences for five other downy mildew species including those affecting mint, sage/rosemary and spinach. This analysis demonstrated that *P. matthiolae* is most closely related to *H. parasitica* (affecting brassicas, oilseed rape) and *H. arabidopsisdis* (pathogenic on *Arabidopsis thaliana)* whereas the remaining species cluster on a separate branch (node) indicating they are less closely related; *P. belbahrii* was more closely related to *P. salivia-officianalis* (Sage/Rosemary DM), than to *P. lamii* (mint DM) and *P. effusa* (Spinach DM).

Sequencing of *P. effusa* and *P. farinosa* f. sp *betae,* and additional samples of *P. belbahrii* and *P. matthiolae* was delayed by Covid-19 impacts on an external company. This has prevented the genome assembly and annotation, and the design of any additional LAMP assays for the Spinach and Sugarbeet downy mildew. The outstanding work and complete datasets shall be released as an Annex to the current report with links to relevant sequence repositories as soon as they become available.

All raw sequencing data will be made publicly available to ensure the greatest access for the research community.

Improved access to enhanced genomics resources generated during the project will greatly improve researchers' abilities to investigate differences in gene content and function underlying host-range, virulence and pathogenicity in important downy mildew species.

Financial Benefits

The new isothermal diagnostic developed for *P. matthiolae* has been shown to be effective for testing seed and plant material for DM infection. The assays could be utilized to screen for the presence of stock DM to help prevent new outbreaks and facilitate early disease detection to help reduce the impact on crops. The assays can be used freely for research purposes, but commercial testing can only be delivered through a licensed service or kit provider (which can require a minimum order before manufacture). LAMP assays kits are designed to be implemented at a low unit cost (<£100 including labor), with minimal operator training and without the requirement for specialized laboratory equipment. Extraction kits enabling preparation of crude templates have facilitated the use of LAMP assays for point of entry and near/in-field screening for plant pathogens. Semi-portable LAMP units retail at approx. £ 9 K whereas fully mobile LAMP units retail at approx. £14.5 k, making them accessible to outreach and extension service providers. Access to improved testing will have direct economic benefits by minimizing losses from unsaleable produce, by reducing crop-destruction to control disease through early detection, and also through a reduction in fungicide treatments.

SCIENCE SECTION

Introduction

Downy mildews caused by various oomycete species are a major constraint on production limiting quality and reducing economic value of high value leafy herbs and ornamentals. Sweet basil, chives, mint, parsley, rosemary, sage and spinach are examples of herb and vegetable crops in which downy mildew outbreaks can have serious negative impacts on productivity. Economic losses for growers from downy mildew outbreaks can be high, necessitating the complete removal of infected crops. *Peronospora destructor* is observed to affect chives, onions and other alliums, *P. belbahrii* affects Sweet basil, Agastache and Coleus[2-5]*, P. lamii* infects mint, *and P. salvia-officinalis*[6] is implicated in causing disease on sage and rosemary, whereas *Plasmopara petroselini* is pathogenic on Parsley. This highlights how a number of different species are responsible for causing downy mildew (DM) diseases on closely-related herbs.

As obligate biotrophs, downy mildew species lack key biosynthetic pathways requiring them to obtain essential nutrients (i.e. sulphur, nitrogen[7]) from their host, and as a result many downy mildews have adapted to become highly specialised on very specific host plants, often with over-lapping host ranges. Physical symptoms often have similar appearance and morphological features in the different species, particularly in spore shape and size. This has historically lead to challenges in accurate species identification and only more recently have molecular genetics strategies been applied to discriminate closely-related downy mildews, facilitating subsequent revision of species names[8,9].

Strategies to control DM are reliant on various practices including the sowing of disease-free seed, crop rotation, selecting uncontaminated production sites (*P. destructor*), the use of resistant varieties where appropriate material is available (i.e. basil, chives) and chemical treatments. Curative spray treatments including Fubol Gold (metalaxyl-M + mancozeb) have been demonstrated to be effective against *P. belbahrii* and *P. destructor* [10,11]. However, the increasing pressure to reduce pesticide inputs, issues with fungicide insensitivity in the target DMs, a risk of exceeding residue limits with reduced efficacy in control [12,13] and decreasing range of chemistries available to growers e.g. impending loss of dimethomorph, necessitates that other means are employed to mitigate against disease.

Provision for practical diagnostics for testing of seed and soil is another major strategy to mitigate against the effects of downy mildew. Molecular diagnostic techniques including qPCR and LAMP have been established as effective methods for accurately detecting downy mildew pathogens, including *P. belbahrii* and *P. effusa* [15-16, 1]. Whilst a powerful tool, unfortunately diagnostic assays are not currently available for many other key downy mildews, limiting their application as a practical management strategy for preventing downy mildew diseases. The development of molecular diagnostics approaches is largely dependent on obtaining accurate genetic information from specific downy mildew species.

Next-generation sequencing technologies now offer researchers a fast, efficient platform for obtaining high-quality genomics data, often directly using their own labs and computing facilities. Current genetic and genomic resources have been developed for model downy mildew *H. arabidopsidis* and certain economically relevant downy mildew species including *P. effusa, P. viticola, P. helstedii, and P humuli* [1,17-19]*.* However, the development of tools and resources that would enable extensive functional and evolutionary genomics studies, and even practical diagnostics tools for downy mildews has received less attention until now, limiting progress. Contemporary sequencing platforms such as those offered by Illumina and Oxford Nanopore Technologies are fast, effective methods to generate high-quality genomics data for conducting a broad range of applications and offer an exciting opportunity to enhance capability for investigating pathogenic downy mildew species.

Improving genomics resources for key downy mildew species would help facilitate fundamental comparative genomics investigations enabling differences in gene content and function underlying host-range, virulence and pathogenicity to be investigated more effectively. This will help contribute towards developing novel, more effective methods of controlling downy mildew pathogens on crops. Moreover, provision of practical diagnostics tools will enable growers to reduce the economic implications caused by downy mildew by minimising new disease outbreaks in production systems.

To this end, the downy mildews selected for the investigation were first cultured using single spore isolation to help reduce potential genetic heterogeneity and then multiplied as pure strains on host plants to provide spore material from which high-molecular weight DNA was then extracted for sequencing. Long-read libraries from *P. belbahrii* and *P. matthiolae* were sequenced on the Oxford Nanopore MinION platform whereas Illumina short-read sequencing was out-sourced to a service provider. Long-read data was subsequently assembled generating initial *de novo* draft assemblies enabling preliminary phylogenetic analysis based in ITS sequences to be conducted and isothermal diagnostic assays design. Unfortunately due to their involvement in the response to the Covid-19 pandemic, the service provider has been unable to complete the work. Outstanding work will be completed as a matter of priority when the situation normalises.

The absence of short-read data has prevented the completion of the *de novo* genome assemblies and is currently limiting the amount of information available for conducting the proposed preliminary phylogenetic studies. This will be addressed as soon as data for gene assembly and annotation becomes available. To this end, the long-read data has been used to identify key house-keeping gene sequences for conducting preliminary phylogenetic investigations and is described herein.

Materials and methods

Culturing downy mildews

Symptomatic plant material was obtained for sweet basil, column stock, red onion, lemon balm/sage, spinach, sugar beet, as potential sources for isolating respective downy mildews, *P. belbahrii, P. matthiolae, P. destructor, P. lamii, P. salvia-officianalis, P. effusa* and *P. farinosa.* Due to the time of year and short-term nature of this project, it was not possible to obtain viable downy mildew samples from parsley, lemon balm, sage or onion/chive. Viable samples that could be successfully transferred onto acceptor-plants for multiplication and subsequently isolated as pure strains were obtained for downy mildews on basil, column stock, spinach and sugar beet. Single-spore isolates for *P. belbahrii, P. effusa, P. farinosa* and *P. matthiolae* were multiplied on their respective host plants cultivated under growth room conditions to ensure purity of cultures (16h day/8h night; Basil: 23°C/18°C day/night; Column stock, Spinach, Sugarbeet 19°C/15°C).

DNA extraction

Fungal spores were collected from fresh, sporulating plants and concentrated using centrifugation at 12 k rpm for 60 sec. Spores were suspended in extraction buffer and homogenized on a genogrinder at 1,000 rpm for 60 sec; DNA for sequencing was extracted using a Qiagen Plant/Fungi DNeasy kit according to the manufacturer's protocol (Qiagen, UK) and analyzed visually on an agarose gel (1 %, run for 1 hour @ 50 v) for integrity and a Qubit 3.0 to provide A260/280, 260/230 values (Life technologies, UK). High-molecular weight DNA templates were prepared for downy mildew species, generating sufficient material for short and long-read sequencing runs.

Nanopore sequencing

Long-read libraries were prepared using 1 µg of HMW DNA and Oxford Nanopore Technology SQK-LSK109 sequencing kits according to the manufacturer's protocol. Libraries were sequenced on a MinION device on an individual single R9.4.1 flow cell for a period of up to 72 hours. Sequencing was monitored and run metrics generated using the ONT MinKnow GUI and base calling implemented with Guppy 3.3.3 (ONT).

Genomes were assembled using Smartdenovo[20]/Wtdgb2[21] and subsequently corrected using Nanopolish. Assembly metrics were generated in QUAST[22], single copy genes were estimated using BUSCO Stramenopiles reference database comprising of 100 independent loci[23].

ITS sequence discovery, alignment and phylogenetic analysis

Newly generated ITS sequences were screened against the NCBI public database using BLAST (Basic Local Alignment Search Tool; https://blast.ncbi.nlm.nih.gov/Blast.cgi) using standard settings. *P. matthiolae* (>utg48754) and *P. belbahrii* (>utg37894) ITS sequences were then aligned and compared with the most closely related species identified from the respective BLAST searches (*H. arabidopsidis, H. parasitica, P. effusa, P. lamii and P. salvia-officianalis)*. ITS sequence for the respective species (Appendix) were aligned and neighbour-joining tree parameters were generated using ClustalW using standard parameters and the trees were visualised as a phylogram using Dendroscope V3.7.2[24].

LAMP assay design

Isothermal amplification assays for *P. matthiolae* were designed on ITS sequences using LAMP designer (Premier Biosoft, USA). Primer sets were synthesized by Sigma-Aldrich (Merck, UK). The *P. belbahrii* assay had previously been validated in EU H2020 project EMPHASIS, using the following reaction conditions.

P. matthiolae LAMP sequence primers:

Pm_F3	TAGTTTGGAGGCTGTCCG
Pm_B3	CGCAGCAATAACACCTGTA
Pm_FIP	TCAGCCGAAGCCAATCATACCCAGTCGGCGATCTGTTTG
Pm_BIP	CTGACGTTTATTGAATATTATTCGGTTGTGGTAGAGAAAGCCAAGTCACACAC
Pm_LoopF	TCAGCCGAAGCCAATCATACC
Pm_LoopB	CTGACGTTTATTGAATATTATTCGGTTGTGGTAGA

Results

Isolate culturing, DNA extraction for sequencing

Pure cultures of *P. belbahrii, P. effusa* (race-type 8), *P. farinosa* f. sp. *betae* and *P. matthiolae* were obtained from infected plant material and maintained under growth room conditions on their respective host plants in order to provide sufficient quantities of spores for DNA extraction (Images 1-4). Conidiophores were washed with distilled water to collect fresh spore material in order to obtain high-quality DNA. Conventional DNA extraction using column-based kits yielded sufficient quantities of template (>1 µg; 50 ng/µl) for Illumina and Nanopore sequencing library preparation.

Images 1-4: Sporulating lesions of 1: *P. belbahrii;* 2: *P. effusa;* 3: *P. farinosa* f. sp *betae;* 4: *P. matthiolae*

Draft long-read assembly and annotation

Assembly sizes from the nanopore sequencing runs were approx. 48 Mb for *P. belbahrii* and 112 Mb for *P. matthiolae*. The *P. belbahrii* assembly is observed to have a fewer number of contigs (238 compared to 551), greater N50 (495690 bp to 425014 bp), largest contigs (2513017 bp to 1475335 bp), greater coverage (130 x compared to 60 x) and greater number of predicted genes (11,065 to 6,290) compared to the *P. matthiolae*. Although seemingly generating a better quality assembly for *P. belbahrii,* the *P. matthiolae* assembly was observed to have greater overall contiguity with a fewer number of scaffolds (14 compared to 61) once the respective genomic contigs were aligned.

The assemblies were assessed for the number of single copy orthologues as a metric of gene content to estimate the level of completeness using the BUSCO installation[23]. Nanopore assemblies for *P. belbahrii* and *P. matthiolae* were observed to have BUSCO scores of 95% and 97%, respectively, based on comparison with the Stramenopiles reference sequence database. This indicated the assemblies were of good quality with relatively high information content.

Long-read assemblies for both species are currently being refined and *de novo* genomes for *P. effusa* and *P. farinosa* will be implemented in the coming weeks. Gene annotation will then be performed across the four species to yield a functional resource for conducting a broad range of genomics applications. The assemblies will be further refined when short-read sequencing data becomes available from the service provider. Raw long-read data generated for the respective samples will be uploaded onto the European Short Read archive (ERA) as soon as it has been completed.

Table 1. Genome assembly metrics for MinION sequencing runs (bp: base pairs)

Genome assembly metrics	*P. belbahrii*	*P. matthiolae*
Total length (bp)	48,428,102	111,957,414
Largest contig (bp)	2513017	1475335
N50 (bp)	495690	425014
L50	25	74
Contigs (bp)	238	551
GC (%)	48.56	48.17
Scaffolds (bp)	31	14
Mean coverage (x)	130	61
BUSCO (%, Straminopiles)	95%	97%
Predicted unique genes	11065	6290

Comparative genomics, phylogenetic analysis, and diagnostics

P. belbahrii and *P. matthiolae* ITS sequences (partial 18 s, ITS 1, 28 s, ITS 2, 5.8 s) were retrieved from their respective long-read assemblies and screened against public sequence databases (NCBI) in order to identify relevant sequences in the most closely related downy mildew species. *P. belbahrii* was observed to be most closely related to *P. salvia-officinalis* (99% similarity over 795 bp) based on ITS sequences, whereas *P. matthiolae* was most closely related to *H. parasitica* (99% similarity over 1568 bp) and *H. arabidopsidis* (98% over 1508 bp).

Alignment of the new ITS sequences with sequences retrieved for *H. parasitica, H. arabidopsidis, P. effusa, P. lamii* and *P. salvia-officinalis* indicated there were high levels of homology between the respective species. Basic neighbor-joining trees were used to visualize the genetic relationships between the closely related downy mildews (Figure 1). *P. belbahrii* and *P. salvia –officianalis* were observed to cluster together indicating they are very closely related, whereas *P. lamii* was placed on an adjacent node indicating this species is more divergent, but still likely to be closely related to basil and sage/rosemary downy mildews. The remaining species occupied two separate clusters on the tree; *P. effusa* and *P. farinosa* f sp. *betae* were observed to cluster together highlighting their similarity, whereas the Brassiciolous downy mildews *H. arabidopsidis, H. parasitica* and *P. matthiolae* formed a separate, independent cluster. *P. matthiolae* appeared to be most closely related to *H. parasitica*. In summary, *P. effusa* and *P. farinosa* are less-closely related to *P. lamii,* and *P. belbahrii / P. salvia-officianalis* based on their ITS sequences, whereas downy mildews affecting column stock, Brassica and *Arabidopsis thaliana* are even more divergent.

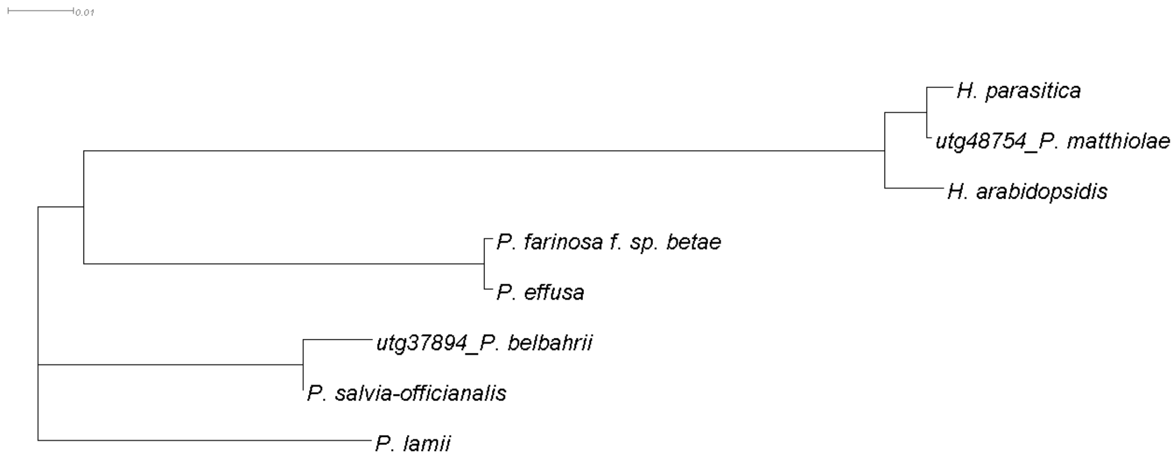

Figure 1. Neighbor joining tree based on ITS sequences from *H. parasitica* (Hp), *H. arabidopsidis* (Hpa), *P. belbahrii* (Pb), *P. effusa* (Pe), *P. farinosa* f. sp. *betae P. lamii* (Pl), *P. matthiolae* (Pm) and *P. salvia-officinalis* (Pso).

LAMP isothermal diagnostic design

A new loop-mediated Isothermal diagnostics assay has been designed for *P. matthiolae* using the *de novo* genome assembly and has been tested for specificity on pathogen and un-infected *Matthiola incana* DNA samples. This has potential for detecting downy mildew infection in cut plant material.

LAMP assays will be designed for *P. effusa* and *P. farinosa* as sequence information for new genome assemblies become available. The three new assays will then be validated to EPPO standards (98/3) using a panel of DNA from target and non-target species in order to demonstrate specificity and reproducibility. The outputs from this testing will be submitted as an Annex to this report.

Discussion

Unfortunately due to the timing of the project occurring after the growing season, viable samples could not be obtained for three of the four target species despite liaising with the British Herb Trade Association, growers, seed suppliers, academic researchers and local horticultural societies; infected plant samples of red onion, lemon balm and sage displaying symptoms of *P. destructor* and suspected *P. lamii* were obtained from growers but unfortunately it was not possible to obtain viable propagules to transfer the infections. No contaminated parsley seed lots or plant samples could be obtained during the project.

Cultures of basil downy mildew were obtained from the NIAB collection and bulked on cv. Sweet Genovese seedlings.

In an effort to account for the lack of samples, the decision was taken to sequence three other closely related downy mildew species affecting high value crops and ornamentals; *P. effusa,* affecting spinach, *P. farinosa* on sugarbeet, and *P. matthiolae* on column stock. The selection was particularly relevant for potentially improving understanding in morphologically similar *P. effusa* and *P. farinosa* affecting the respective related host-species spinach (*Spinacia oleracea*) and sugarbeet (*Beta vulgaris*) which are both cultivated members of the *Amaranthaceae.* Until recently, spinach downy mildew was named *P. farinosa* f. sp. *effusa* highlighting both the similarity with sugarbeet downy mildew and the difficulty in distinguishing between downy mildew species on morphological parameters alone[25.]This is a similar situation to that reported for *P. belbahrii*, where basil downy mildew, also affecting coleus and agastache, was often misidentified as *P. lamii,* which itself affects a number of the Lamiaceae[3.]Single spore isolation was used as a strategy to reduce potential genetic heterogeneity associated with potentially mixed downy mildew populations. Although time consuming, it was possible to multiply sufficient quantities of spores from individual strains and extract DNA using conventional extraction kits yielding sequencing-grade templates for *P. belbahrii* and *P. matthiolae* (A260:230:>1.8;A260:280:>1.8).

Nanopore libraries were prepared and sequenced in-house providing a fast, effective platform for sequencing new pathogen strains. Nanopore sequencing, as with other long-read strategies is observed to have a relatively high error rate (~15%) compared to Illumina's reported 0.1%, and thus necessitates subsequent rounds of polishing to remove errors. This process can be made even more robust by incorporating short-read data to check the long-read consensus against, enabling users to generate contiguous genome sequences with high confidence. Unfortunately our Illumina sequencing service provider experienced severe disruptions due to the Covid-19 epidemic leading to delays in the completion of the short-read libraries. This data will be analyzed and incorporated into the long-read datasets as it becomes available.

Despite the lack of short-read datasets, the primary iterations of the *de novo P. belbahrii/P. matthiolae* long-read genomes were still very promising. Assembly sizes from the nanopore sequencing runs generated were approx. 48 Mb for *P. belbahrii* and 112 Mb for *P. matthiolae*. Recently sequenced *P. belbahrii* genomes have reported sizes ranging from between 35-59.2 Mb[26,27]; our first preliminary draft long-read genome falling somewhere in the middle of these at 48 Mb. Subsequent iterations will likely reduce the length further. The new long-read *P. belbahrii* assembly has doubled the N50 value and reduced the L50 value by 1/3 demonstrating greater overall contiguity, as was anticipated using this type of sequencing approach. Duplication was identified to be a potential factor responsible for the large difference in size between the two established *P. belbahrii* assemblies and it will be interesting to assess the effects of removing duplicated sequences and potential contaminants on the size of the final assemblies across the four target species.

The use of pure strains should help to account for some of the genetic heterogeneity observed in the previous studies, however downy mildew conidia are often multi-nucleate so absolute homogeneity cannot be guaranteed without separating and sequencing individual nuclei. More insight will be gained into levels of potential contamination, duplication, repetitive DNA and genetic heterogeneity when the assembly filtering and drafting process has been completed.

The *P. matthiolae* (112 Mb) genome assembly was observed to be twice the length of *P. belbahrii* (48 Mb). It was anticipated that *P. matthiolae* may have a larger genome as the closely related species *H. arabidopsidis* exhibits a genome of approx. 77.8-81.6 Mb[7]. The *P. belbahrii* genome is reported to comprise of only 16% repetitive DNA[26], whereas *H. arabidopsidis* contains 42% so it is reasonable to postulate a similar case for the closely-related *P. matthiolae.* Repetitive DNA such as transposable elements and tandem repeats are a common feature of expanded genomes and could be responsible for some of the additional length. More insight will be gained into levels of potential contamination, duplication, repetitive DNA and genetic heterogeneity when the assembly filtering and drafting process has been completed.

The greater coverage and higher quality assembly generated for the *P. belbahrii* is suggested to relate to the higher base-pair count of the potentially expanded genome of *P. matthiolae* and the higher quality of DNA that can be obtained for basil downy mildew*. P. matthiolae* conidia are more difficult to isolate from leaf tissue due to a waxy composition, and subsequently the nucleic acid extracted is usually of lower quality due to small quantities of contaminants being carried over. Sequencing library preparation requires high purity DNA to generate the best quality data and therefore even small amounts of inhibitory compounds can have negative effects on efficiency.

The *P. belbahrii* genome assembly could be resolved into fewer contigs compared to *P. matthiolae;* this likely reflects the expansion in the size of the *P. matthiolae* genome but also potentially issues with resolving repetitive regions of the genome such as homopolymers, transposable elements, tandem repeats that could be present. However, despite higher overall quality the *P. belbahrii* assembly, the *P. matthiolae* could be resolved into fewer scaffolds, indicating greater overall contiguity.

BUSCO scores of 97% and 95 % were obtained for *P. matthiolae* and *P. belbahrii*, respectively, with the high levels of recovery indicating that the new assemblies were of good quality and had captured sufficient sequence information represented within the respective genomes despite only screening against a relatively small set of reference genes. Despite similar BUSCO scores, nearly half the number of unique genes were reported for *P. matthiolae* compared to *P. belbahrii* reflecting the preliminary iteration of the *de novo* assembly. This will undoubtedly be improved upon with further iterations and integration of short-read data. Moreover, it should be stressed that the draft assemblies were generated to represent a preliminary representation of *P. belbahrii* and *P. matthiolae* genomes, and will be improved upon with further bioinformatics polishing of the long-read data and through the addition of short-read Illumina data.

The newly drafted genomes were screened for ITS sequence in order to provide targets for conducting phylogenetic analysis and design specific isothermal diagnostic assays. The phylogenetic analysis indicated that *P. belbahrii* was more closely related to *P. salvia-officianalis* than *P. lamii*, demonstrating the power of molecular approaches for distinguishing between morphologically similar species compared to problematic visual identification. *P. effusa* and *P. farinosa* f sp. *betae* were observed to cluster together during the phylogenetic analysis, as was anticipated for such closely related (and previously synonymous species).

In contrast, *P. matthiolae* appeared to be most closely related to *H. parasitica* and *H. arabidopsidis,* with high levels of homology observed in their ITS sequences (>95%).

P. matthiolae, *H. arabidopsidis* and *H. parasitica's* respective hosts *Matthiola incana, Arabidopsis thaliana* and *Brassica* spp. are all members of the Brassicaceae, demonstrating how closely related downy mildew species have evolved to become specialized on similar/related host species[28]. Interestingly, *H. parasitica* has been implicated in causing downy mildew on column stock in the USA[29,30] highlighting the case for over-lapping host ranges and complex speciation in may related downy mildew species. It will be interesting to compare the respective host-ranges of the three closely-related Brassiciolous downy mildews, particularly *P. matthiolae* in order to assess if it displays any affinity for colonizing *Arabidopsis,* or if it exhibits a discrete host-range as do *H. arabidopsidis* and *H. parasitica*.

It should be stressed that this has only been a preliminary analysis based on genetic variation in a single reference gene and one isolate of each downy mildew species, and therefore it is important to avoid over-interpretation of the relationships. A more thorough phylogenetic analyses, based on non-synonymous changes in protein coding genes will be used as a strategy to compare genetic diversity between the four downy mildews and additional species.

It will be interesting to investigate the genetic relationship between *P. matthiolae* and its close relatives to establish if reclassification to a *Hyaloperonospora* would be more appropriate.

Conclusions

A novel isothermal LAMP assay has been designed for detecting *P. matthiolae* on column stocks based on newly sequenced ITS regions. This has been compared to closely related downy mildew and related species in an effort to design a highly specific diagnostic and will now be tested against a panel of related organisms to ensure efficacy, in conjunction with new LAMP assays for *P. effusa* and *P. farinosa.* New isothermal diagnostics will provide growers and seed producers with enhanced tools for limiting the impact of disease in production environments.

Draft long-read assemblies have been generated for two key downy mildew pathogens on brassica and sweet basil hosts; genome sequencing of spinach and sugarbeet downy mildews is currently being finalized. Additional Illumina short reads shall be utilized for polishing the respective genomes and generating hybrid assemblies. The *de novo* sequencing of *P. matthiolae* has enabled us to identify high-levels of homology with *H. arabidopsidis,* positioning the pathogen alongside the other Brassiciolous downy mildews, and providing essential resources for conducting comparative genomics studies in closely-related downy mildew.

Glossary

Assembly: A series of ordered nucleotide sequences representing an organism's genetic content, usually constructed using a series of software tools

Contigs: Contiguous sequence of DNA that make up a genome assembly; contigs are comprised of aligned sequence reads

LAMP: Loop mediated isothermal AMPlification

qPCR: Quantitative Polymerase Chain Reaction

References

1. Fletcher et al., 2018. Comparative genomics of downy mildews reveals potential adaptations to biotrophy. BMC Genomics 19: 851
2. Belbahrii et al., 2005. Phylogenetic analysis and Real Time PCR detection of a presumably undescribed *Peronospora* species on sweet basil and sage. Mycological Research 109, 1276-1287
3. Thines et al., 2009. Identity of the downy mildew pathogens of basil, coleus, and sage with implications for quarantine measures Mycological Research 113:5 532-540

4. Henricot et al., 2010. Peronospora belbahrii causing downy mildew disease on Agastache in the UK: a new host and location for the pathogen Plant Pathology 59:4 801
5. Denton et al., 2015. First record of downy mildew, caused by Peronospora belbahrii , on Solenostemon scutellarioides in the UK. New Disease Reports 31:14, http://dx.doi.org/10.5197/j.2044-0588.2015.031.014
6. Humphrey-Jones et al., 2008. First report of the downy mildew Peronospora lamii on Salvia officinalis and Rosmarinus officinalis in the UK. Plant pathology 57:2 372
7. Baxter et al., 2010. Signatures of adaptation to obligate biotrophy in the Hyaloperonospora arabidopsidis genome.10:330, 1549-1551. doi: 10.1126/science.1195203
8. Reithmüller et al., 2002. Phylogenetic relationships of the downy mildews (Peronosporales) and related groups based on nuclear large subunit ribosomal DNA sequences. Mycologia 94:5 834-849 https://doi.org/10.1080/15572536.2003.11833177
9. Hoffmeister et al., 2020. Two new species of the Peronospora belbahrii species complex, Pe. choii sp. nov. and Pe. salviae-pratensis sp. nov., and a new host for Pe. salviae-officinalis. Fungal Systematics and Evolution. 10.3114/fuse.2020.06.03.
10. AHDB Final Report PE024a, 2017. Basil: Improving knowledge and Control of Downy Mildew in protected and outdoor crops
11. AHDB Report CP189c, 2006. New Products for the Control of Downy Mildew (Peronospora destructor) in Bulb and Salad Onions
12. Cohen et al., 2013. First Report of the Occurrence and Resistance to Mefenoxam of Peronospora belbahrii, Causal Agent of Downy Mildew of Basil (Ocimum basilicum) in Israel. Plant disease 97(5):692 doi: 10.1094/PDIS-12-12-1126-PDN.
13. Gilardi, 2019; pers. comm.
14. Shao D, Tian M, 2018. A qPCR approach to quantify the growth of basil downy mildew pathogen Peronospora belbahrii during infection. Current Plant Biology, DOI: 10.1016/j.cpb.2018.09.003
15. H2020 Emphasis project: http://www.emphasisproject.eu/; https://cordis.europa.eu/project/rcn/193276/factsheet/en.
16. Feng et al., 2014. Multiplex real-time PCR assays for detection of four seedborne spinach pathogens. Journal of Applied Microbiology 117, 472-484
17. Yin et al., 2017. Genome sequence of Plasmopara viticola and insight into the pathogenic mechanism. Scientific Reports 7: 46553
18. Sharma et al., 2015. Genome analyses of the sunflower pathogen Plasmopara halstedii provide insights into effector evolution in downy mildews and Phytophthora. BMC Genomics 16:741. 10.1186/s12864-015-1904-7.
19. Rahman et al., 2019. Genome Sequencing and Transcriptome Analysis of the Hop Downy Mildew Pathogen Pseudoperonospora humuli Reveal Species-Specific Genes for Molecular Detection. Phytopathology 109:8 https://doi.org/10.1094/PHYTO-11-18-0431-R
20. Smartdenovo: https://omictools.com/smartdenovo-tool
21. Wtdgb2: https://github.com/ruanjue/wtdbg2
22. Gurevich et al., 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013 Apr 15;29(8):1072-5. doi: 10.1093/bioinformatics/btt086.
23. Simao et al., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31(19):3210-2 doi: 10.1093/bioinformatics/btv351.
24. Huson et al., 2007. Dendroscope: An interactive viewer for large phylogenetic trees. BMC Bioinformatics 8: 460

25. Choi et al., 2007. Re-consideration of Peronospora farinosa infecting Spinacia oleracea as distinct species, Peronospora effusa. Mycological Research 111:4 381-391
26. Thines et al., 2019. The genome of Peronospora belbahrii reveals high heterozygosity, a low number of canonical effectors and CT-rich promoters. BioRxiv doi: https://doi.org/10.1101/721027
27. Jackman et al., 2017 ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter. Genome Res. 27:768-777.
28. Thines & Choi, 2016. Evolution, Diversity, and Taxonomy of the Peronosporaceae, with Focus on the Genus Peronospora. Phytopathology Review 106:1 https://doi.org/10.1094/PHYTO-05-15-0127-RVW
29. Koike et al., 2000. Downy Mildew of Stock, Caused by Peronospora parasitica, in California. Plant Disease 84:1. https://doi.org/10.1094/PDIS.2000.84.1.103E
30. Koike, 2018. Diseases of Stock. Handbook of Florists' Crops Diseases. Springer https://link.springer.com/referencework/10.1007/978-3-319-39670-5

Appendix: Downy Mildew ITS sequences

>utg37894_Pb

AGGATCATTACCACACCTAAAAACTTTCCACGTGAACCGTATC-
AACCCAATAATTTGGGGGTTAATTTGGTAATGGCTGCTGGCATCTTTTTGTTGGCTGGC
TGTTACTGAGTAAGCCCTATCATGGCGAGAGTTTGGACCTTTGTCTGAACAAGTAGCTA
CGATTGTAAACCCTTTCTTAATTACTGAATATACTGTGGGGACGAAAGTCTCTGCTTTTA
ACTAGATAACAACTTTCAGCAGTGGATGTCTAGGCTCGCACATCGATGAAGAACGCTG
CGAACTGCGATACGTAATGCGAATTGCAGGATTCAGTGAGTCATCGAAATTTTGAACG
CATATTGCACTTCCGGGTTAGTCCTGGAAGTATGCCTGTATCAGTGTCCGTACAACAAA
CTTGGTTTTCTTCTTTCCGTGTAGTCGGTGGAGGATATGCCAGATATGAAGTGTCTTGC
GATTGATTTTT-GTATCGGTTGCGAGTCCTTTTAAACGTAAAACACTGTACTTCTCTTT-
GCTCGAAAAGCGTGGCGTTGCTGGTTGTGAAGGCTGTCCATGTGACCAGTCGGCGAT
CGTTTTGTCTACTATGGCGTTTAATGGAGAAGTATTTGATTCGCGGTATGATTGGCTTC
GGCTGAACAGGCGCTGATTGAATATTTTTCCTGCTATGGCGGTATGAACTTGATGAAC
CGTAGTCGTATGTTGACTTGGCTTTTGAATTGGCTTTGCTGTTGCGAAGTAGAGTGACA
GTTTCGGCTGTCGAGGGTCGACCGATTTGGGAAATTTGTGTGATGCGACTTCGGTTGC
GTTGCATCTCAACTCGACCTGATATCAGGCAAGATTACCCGCTGAACTTA-

>EU049268.1_Hp

GGATTCATTACCACACCTAAAAACTTTCCACGTGAACCGTATCAACAACATAAATTGGG
GGTGATCTTGACGGTAGCTGCTGGTTTTTTACTGGCGGTAACTATTGAGTGAACCTCAT
CATGACGAACCGTTGGGCTTTTGTCTGGCGTAGTAGAATAAATTTTAAACCCTTACTTA
ATTACTGATTATACTGTGGGGACGAAAGTCTCTGCTTTTATTCTAGATAACAACTTTCAG
CAGTGGATGTCTAGGCTCGCACATCGATGAAGAACGCTGCGAATTGCGATACGTAATG
CGAATTGCTAAATTCAGTGAGTCATCGAAATTTTGAACGCATATTGCACTTCCGGGGCT
AGTCCTCAGAAGTATGCCTCTATCAGTGTCCACACAACAAACTTGGTTTTCTTCTTTCC
GTGTATTCGGTGGAGGATATGCCAGATATGAAGTGTCTTGCGGCTGGTTTTCGAATCG
GCTGCGAGTCCTTTTAAATAAATAGAACTGTACTTCTCGTGCTCGAAAAGCGTAACGAT
ACTAGTTTGGAGGCTGTCCGCGTGGCCAGTCGGCGATCTGTTTGTTAACTGTGACTAT
ATGAAGAAGTATTTGATTTTCGGTATGATTGGCTTCGGCTGAACTGACGTTTATTGGAT
ATTATTCGGTTGTGGTAGACGGATGGGTGAACCGTAGTAGCGTGTGTGACTTGGCTTT
CTTACAGGTGTTATTGCTGCGAAGTAAGGCGTCAACTTCGGTTGTCCGAGGGTCGACC
TATTTGGGAATAGTGTTGTAAAGCTTCGGCTGCACTTCACCTCAAACGGACCTGATATT
AGGCAAGATTACCCGCTGAACTTAAGCATATTAATAA

>EU049263.1_Hpa

>utg48754_Pm

CGGAAGGATC-ATTACCACACCTAAAAA-CTTTCCACGTGAACCGTATCAACAACATAAA

TTGGGGGTGATCTTGACGGTAGCTGCTGGT---TTTTTACTGGC-GGTAACTATTGAGTG

AACCTCATCATGACGAACCGTTGGGCTTTTGTCTGGCGTAGTAGAATAAATTTTAAACC
CTTACTTAATTACTGATTATACTGTGGGGACGAAAGTCTCTGCTTTTATTCTAGATAACA
ACTTTCAGCAGTGGATGTCTAGGCTCGCACATCGATGAAGAACGCTGCGAATTGCGAT
ACGTAATGCGAATTGCTAAATTCAGTGAGTCATCGAAATTTTGAACGCATATTGCACTT
CCGGGGCTAGTCCTCAGAAGTATGCCTTTATCAGTGTCCACACAACAAACTTGGTTTTC
TTCTTTCCGTGTATTCGGTGGAGGATATGCCAGATATGAAGTGTCTTGCGGCTGGTTTT
CGAATCGGCTGCGAGTCCTTTTAAATAAATAGAACTGTACTTCTCGT-
GCTCGAAAAGCGTAACGATACTAGTT-TGGAGGCTGTCCGCG-
TGGCCAGTCGGCGATCTGTTTGTTAACTGTGACTAT----
ATGAAGAAGTATTTGATTTTCGGTATGATTGGC--TTCGGCTGAACTGACGTTTA

TTGGATATTATTCG-GTTGTGGTAG-ACGGA-TGGGTGAACCGTAGTAGCGTGTGTGACT

TGGCTTTCTTACAGGTGTTATTGCTGCGAAGTAAGGCGTCAACTTCGGTTGTCCGAGG
GTCGACCTATTTGGGAA--
TAGTGTTGTAAAGCTTCGGCTGCACTTCACCTCAAACGGACCTGATATTAGGCAAGATT
ACCCGCTGAACTTAAGCATATTAATAAGCGGAGG-

>MN450331.1_Pso

TTAGATGTTCTGGGCCGCACGCGCGCTACACTGATGCGTTCAACGAGTATACAACCTT
GATCGATAGGTCTGGGTAATCTTGTAAATGCGCATCGTGCTAGGGATAGACTGTTGCA
ATTTTCAGTCTTGAACGAGGAATTCCTAGTAAACGCAAGTCATCAGCTTGCATTGATTA
CGTCCCTGCCCTTTGTACACACCGCCCGTCGCACCTACCGATTGAATGACTCGGTGAA
AAATTGAGACTGTTAGTCCGCTTGCTTCATTGCGAGTGAATTGATGGGAACTTTTTTAA
ACCTCGCCATTTAGAGGAAGGTGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGG
AAGGATCATTACCACACCTAAAAACTTTCCACGTGAACCGTATCAACCCAATAATTTGG
GGGTTAATTTGGTAATGGCTGCTGGCATCTTTTTGTTGGCTGGCTGTTACTGAGTAAGC
CCTATCATGGCGAGAGTTTGGACCTTTGTCTGAACAAGTAGCTACGATTGTAAACCCTT
TCTTAATTACTGAATATACTGTGGGGACGAAAGTCTCTGCTTTTAACTAGATAACAACTT
TCAGCAGTGGATGTCTAGGCTCGCACATCGATGAAGAACGCTGCGAACTGCGATACGT
AATGCGAATTGCAGGATTCAGTGAGTCATCGAAATTTTGAACGCATATTGCAYTTCCGG
GTTAGTCCTGGAAGTATGCCTGTATCAGTGTCCGTACAACAAACTTGGTTTTCTTCTTT
CCGTGTAGTCGGTGGAGGATATGCCAGATATGAAGTGTCTTGCGATTGATTTTTGTATC
GGTTGCGAGTCCTTTTAAACGTAAAACACTGTACTTCTCTTTGCTCGAAAAGCGTGGCG
TTGCTGATTGTGAAGGCTGTCCATGTGACCAGTCGGCGATCGTTTTGTCTACTATGGC
GTTTAATGGAGAAGTATTTGATTCGCGGTATGATTGGCTTCGGCTGAACAGGCGCTTAT

TGAATGTTTTTCCTGCTATGGCGGTATGAACTTGATGAACCGTAGTCGTATGTTGACTT
GGCTTTTGAATTGGCTTTGCTGTTGCGAAGTAGAGTGACAGCTTCGGCTGTCGAGGGT
CGACCGATTTGGGAAATTTGTGTGATGCGAC

>MN450325.1_PI

CCGCACGCGCGCTACACTGATGCGTTCAGCGAGTATATAACCTTGATCGATAGGTCTG
GGTAATCTTGTAAATGCGCATCGTGCTAGGGATAGACTGTTGCAATTTTCAGTCTTGAA
CGAGGAATTCCTAGTAAACGCAAGTCATCAGCTTGCATTGATTACGTCCCTGCCCTTTG
TACACACCGCCCGTCGCACCTACCGATTGAATGACTCGGTGAAAAATTGGGACTGTGA
GTCCGCTTACTTCATTGCGAGTGGATTGATGGGAACTTTTTTAAACCTCGCCATTTAGA
GGAAGGTGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCATTACCAC
ACCTAAAAACTTTTCACGTGAACCGTATCAAACACTTTATTGGGGGCTGACTTGGTGGC
GGCTGCTGGTATCTATTTTTTTACTGACTGGCTGCTGCTTTGAGAAAGCCCTATCATGG
CGATCGCTTGGACTTATATTCGAGCTAGTAGCTACTATTTTAAACCCATTCTTAATTACT
GAATATACTGTGGGGACGAAAGTCTCTGCTTTTAACTAGATAACAACTTTCAGCAGTGG
ATGTCTAGGCTCGCACCTCGATGAAGAACGCTGCGAACTGCGATACGTAATGCGAATT
GCAGGATTCAGTGAGTCATCGAAATTTTGAACGCATATTGCACTTCCGGGTTCGTCCT
GGAAGTATGCTTGTATCAGTGTCCGTACATCAAACTTGGTTTTCTTCTTTCCGTGTAGT
CGGTGGAGGATATGCCAGATGTGAAGTGTCTTGCGGCGGCTGATTTTCGGATCTGCT
GCGAGTCCTTTGAAATGTACAGAACTGACTTCTCTTTGCTCGAAAAGTGTGGCATTACT
GGTTGTAAAGGCTGTCCGTGTGACCAGTCGGCGATCGGTTTGTCTGCTGCGGCGTTA
ATGGAAGGATGTTTGATTTGCGGTATGGTTGGCTTCGGCTGAACAGACGCTTATTGGA
CGTACTTCTTGCTGTGGCGGTATGAACTGATGTACCGTAGTTGTATGTGACTTGGCTTT
TTGAATTGGTGTTGCTGTGCGAAGTAGAGTGACAGTTTCGGCTGTCGAGTGTCGACCA

>EU591713.1_Pf_betae

CCACACCTAAAAAACTTTCCACGTGAACCGTATCAACCCAATAATTTGGGGGTTTAATT
GGCGGCTGCTGCTGGTATCTTTTTACTGGCTGGCGGCTGCTGAGCGAACCCTATCGT
GAGCGTTCTGACCTTGGTTGGAGCTAGTAACTTAAATTTTAAACCCATTCTTAAATACTG
AATATACTGTAGGGACGAAAGTCTCTGCTTTTAACTAGATAGCAACTTTCAGCAGTGGA
TGTCTAGGCTCGCACCTCGATGAAGAACGCTGCGAACTGCGATACGTAATGCGAATTG
CAGGATTCAGTGAGTCATCGAAATTTTGAACGCATATTGCACTTCCGGGTTATCCCTGG
AAGTATGCCTGTATCAGTGTCCGTACATTAAACTTGGTTTTCTTCTTTCCGTGTAGTCG
GTGGAGGATATGCCAGATATGAAGTGTCTTGTGGCTGGTTTTCGAATCGGCTGTGAGT
CCTTTGAAATGTATAGAACTGATTCTCTTTGTTCGAAAAGCGTGGCGTTGCTGGTTGTG
AAGGCTGTCAGTGTGACTAGTCGGCGACCGGTTTGTCTGCTATGGCATAAATGGAGGA
GTGTTCGATTCGCGGTATGATTGGCTTCGGCTGAACAGGCGCTTATTAGACGTTCTTC

CTGCTGTGGCGGTATGACCTGGTGAACCGTAGTTCATGCACGACTTGGCTTTTGAATT
GGCTTTGCTGTGTGAAGTAGAGTGACAGTTTCGGCTGTCGAGGATCGACCCATTTGGG
AAATTGTGCTGTGCGACTTCGGTTGCGTGGCATCTCAA

>MG793197.1_Pe

CCACACCTAAAAAACTTTCCACGTGAACCGTATCAACCCAATAATTTGGGGGTTTAATT
GGCGGCTGCTGCTGGTATCTTTTTACTGGCTGGCGGCTGCTGAGCGAACCCTATCGT
GAGCGTTCTGACCTTGGTTGGAGCTAGTAACTTAAATTTTAAACCCATTCTTAAATACTG
AATATACTGTAGGGACGAAAGTCTCTGCTTTTAACTAGATAGCAACTTTCAGCAGTGGA
TGTCTAGGCTCGCACATCGATGAAGAACGCTGCGAACTGCGATACGTAATGCGAATTG
CAGGATTCAGTGAGTCATCGAAATTTTGAACGCATATTGCACTTCCGGGTTATCCCTGG
AAGTATGCCTGTATCAGTGTCCGTACATTAAACTTGGTTTTCTTCTTTCCGTGTAGTCG
GTGGAGGATATGCCAGATATGAAGTGTCTTGTGGCTGGTTTTCGAATCGGCTGTGAGT
CCTTTGAAATGTATAGAACTGATTCTCTTTGTTCGAAAAGCGTGGCGTTGCTGGTTGTG
AAGGCTGTCAGTGTGACTAGTCGGCGACCGGTTTGTCTGCTATGGCATAAATGGAGGA
GTGTTCGATTCGCGGTATGATTGGCTTCGGCTGAACAGGCGCTTATTAGACGTTCTTC
CTGCTGTGGCGGTATGACCTGGTGAACCGTAGTTCATGCACGACTTGGCTTTTGAATT
GGCTTTGCTGTGTGAAGTAGAGTGACAGTTTCGGCTGTCGAGGATCGACCCATTTGGG
AAATTGTGCTGTGCGACTTCGGTTGTGTGGCATCTCAA