

RESEARCH ARTICLE

Open Access

# The genome sequence of the biocontrol fungus *Metarhizium anisopliae* and comparative genomics of *Metarhizium* species

Julie A Pattemore<sup>1\*</sup>, James K Hane<sup>2,3</sup>, Angela H Williams<sup>2,3</sup>, Bree AL Wilson<sup>1</sup>, Ben J Stodart<sup>1</sup> and Gavin J Ash<sup>1</sup>

## Abstract

**Background:** *Metarhizium anisopliae* is an important fungal biocontrol agent of insect pests of agricultural crops. Genomics can aid the successful commercialization of biopesticides by identification of key genes differentiating closely related species, selection of virulent microbial isolates which are amenable to industrial scale production and formulation and through the reduction of phenotypic variability. The genome of *Metarhizium* isolate ARSEF23 was recently published as a model for *M. anisopliae*, however phylogenetic analysis has since re-classified this isolate as *M. robertsii*. We present a new annotated genome sequence of *M. anisopliae* (isolate Ma69) and whole genome comparison to *M. robertsii* (ARSEF23) and *M. acridum* (CQMa 102).

**Results:** Whole genome analysis of *M. anisopliae* indicates significant macrosynteny with *M. robertsii* but with some large genomic inversions. In comparison to *M. acridum*, the genome of *M. anisopliae* shares lower sequence homology. While alignments overall are co-linear, the genome of *M. acridum* is not contiguous enough to conclusively observe macrosynteny. Mating type gene analysis revealed both *MAT1-1* and *MAT1-2* genes present in *M. anisopliae* suggesting putative homothallism, despite having no known teleomorph, in contrast with the putatively heterothallic *M. acridum* isolate CQMa 102 (*MAT1-2*) and *M. robertsii* isolate ARSEF23 (altered *MAT1-1*). Repetitive DNA and RIP analysis revealed *M. acridum* to have twice the repetitive content of the other two species and *M. anisopliae* to be five times more RIP affected than *M. robertsii*. We also present an initial bioinformatic survey of candidate pathogenicity genes in *M. anisopliae*.

**Conclusions:** The annotated genome of *M. anisopliae* is an important resource for the identification of virulence genes specific to *M. anisopliae* and development of species- and strain- specific assays. New insight into the possibility of homothallism and RIP affectedness has important implications for the development of *M. anisopliae* as a biopesticide as it may indicate the potential for greater inherent diversity in this species than the other species. This could present opportunities to select isolates with unique combinations of pathogenicity factors, or it may point to instability in the species, a negative attribute in a biopesticide.

**Keywords:** *Metarhizium anisopliae*, Genome, Pathogenicity, Motifs, Mating-type, Comparative, Genomics

## Background

*Metarhizium anisopliae* is a globally distributed, entomopathogenic fungus that infects many important crop pests including aphids, scarabaeoid beetle larvae and western flower thrips [1-4] (Figure 1). The species was one of the first to be investigated for its use as a biological

control agent and advances in the understanding of its biology and ecology have led to improved biocontrol applications [5]. *M. anisopliae* is regarded as asexual as no teleomorph has been observed [6]. In cases such as these, phylogenetic species boundaries are often used to taxonomically characterize anamorphic fungi [7] and *M. anisopliae* has been well characterized in this regard [5,6,8,9]. Analysis of mating type loci (idiomorphs) however, can enhance our understanding of the genetic mechanisms behind sexual or asexual lifestyles and the potential pathways of genetic exchange.

\* Correspondence: jpattemore@csu.edu.au

<sup>1</sup>Graham Centre for Agricultural Innovation, School of Agricultural and Wine Sciences, Charles Sturt University, Locked Bag 588, Wagga Wagga 2650, NSW, Australia

Full list of author information is available at the end of the article

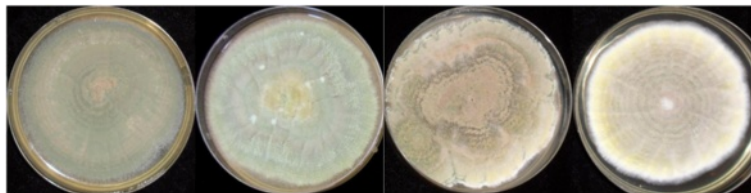


**Figure 1** Green conidia of *Metarhizium anisopliae* isolate Ma69 erupting from the cadaver of a rose-grain aphid (*Metopolophium dirhodum*).

The *Metarhizium* species complex is diverse, including generalist species with a broad host range and specialist species with narrow host ranges. Furthermore, individual isolates can also exhibit a range of cultural variability. In our laboratory, variability of cultures is minimized through the use of single spore isolates and mother cultures from long term storage. Despite best practice, significant variability arises in the colour and amount of sporulation from replicates of the same single spore isolates (BRIP 53293) cultured on SDAY plates of identical composition and grown under identical temperature regimes (Figure 2). Two main morphologies have been observed: 1) highly sporulating olive green cultures and 2) low sporulating tri-colour cultures, that is orange, pale green and olive green. In addition to these morphologies, some cultures also exhibit more abundant fluffy mycelial growth while others tend to sector, a sign of aging [10]. Culture degeneration has been shown to affect the stability of enzyme production (e.g. cuticle degrading enzymes) and secondary metabolite production (e.g. destruxins) and results in extensive downstream gene regulation [10,11].

While possible mechanisms responsible for sectoring may include changes due to physiological and environmental adaptations or differences in transposable elements [11-13], our observations of cultural phenotype variation under laboratory conditions are not well understood. Phenotypic variability which affects sporulation in the laboratory could also affect the multiplication of blastospores in the host haemocoel, isolate pathogenicity, as well as commercial conidial production, which has important implications for maintenance, preservation and the selection of isolates for commercialization [11,14]. The existence of unique divergent gene sets and expansion of gene families within *M. anisopliae* may provide an important genetic resource for the continued development of this important entomopathogen as a biopesticide.

The recent publication of reference genomes and comparative genomics of the generalist *M. anisopliae* and the locust-specific pathogen *M. acridum* has further enhanced our understanding of the biology of these entomopathogenic fungi and the molecular basis of host-specificity [5,13]. However the isolate sequenced as the



**Figure 2** Examples of cultural variability found among single spored isolates from the parent culture of *M. anisopliae* BRIP 53293, grown on SDAY plates of identical composition under identical temperature regimes. The main morphologies observed are (from left to right) one highly sporulating olive green culture and low sporulating tri-colour cultures.

reference strain for *M. anisopliae* (ARSEF23) was subsequently re-classified to *M. robertsii* [13], a phylogenetically distinct species within the *Metarhizium* PARB clade (*M. pingshaense*, *M. anisopliae*, *M. robertsii* and *M. brunneum*) [5].

In light of the well-supported divergence of ARSEF23 from the main *M. anisopliae* clade [5,9] and re-classification to *M. robertsii*, the primary aim of this study was to assemble an accurate, annotated, draft reference genome for *M. anisopliae* using our isolate Ma69. We used a comparative genomics approach, not to repeat the excellent work of the previous publication [13], but to complement it, add to the body of knowledge and establish new genomic resources for the *Metarhizium* genus. In this study, we compare the genomes of *M. anisopliae*, *M. robertsii* and *M. acridum* to identify the key suite of genes which differentiate *M. anisopliae* from the other two species. We used a multi-faceted bioinformatics approach to identify genes that were divergent between the three *Metarhizium* species and to assign putative functions to them where possible. We identify a suite of effector-like genes that are predicted to be specific to *M. anisopliae*, significant differences in the repetitive DNA complements, repeat-induced point mutations and mating type gene composition between the three species and discuss the implications of these findings.

## Results

### Phylogenetic validation, genome sequencing and assembly of *M. anisopliae* isolate Ma69

The rDNA internal transcribed spacer (ITS) region of *M. anisopliae* isolate Ma69 was sequenced and analyzed by BLASTN to confirm its identity [10]. One hundred hits to *M. anisopliae* ITS sequences with an e-value of 0.00 were obtained, confirming isolate Ma69 as *M. anisopliae* (Additional file 1. Top ten hits only).

The genome of isolate Ma69 was then shotgun-sequenced by generating two Illumina libraries, a 100 bp paired-end library and a 3 kb mate-paired library, comprising 5.61 Gb and 9.29 Gb of raw data respectively. This corresponded to approximately 380X coverage of the final genome assembly. A total of 142.9 million reads (96.02%) were retained for assembly after quality-control filters were applied for the removal of adapter sequence and regions of low base-call quality or low sequence complexity. Initial *de novo* assembly using only paired-end reads produced an assembly containing 1,567 scaffolds with an N50 of 138 and an N50 length of 85,355 bp. Contigs were scaffolded with the 3 kb mate-pair library, resulting in 577 scaffolds with an N50 of 11 and an N50 length of 1,243,138 bp. The genome assembly statistics and general features were tabled along with those of *M. robertsii* and *M. acridum* [13] (Table 1). The

**Table 1 Summary of genome and gene content for *Metarhizium* isolates, *M. anisopliae* Ma69, *M. robertsii* ARSEF23 and *M. acridum* CQMa102**

Features	<i>M. anisopliae</i> Ma69	<i>M. robertsii</i> ARSEF23	<i>M. acridum</i> CQMa102
Genome assembly size (Mb)	38.6	39.04	38.05
Genome assembly coverage	380x	100x	107x
Genome assembly N50	11	7	36
Genome assembly N50 length	1,243,138	1,958,674	329,480
Total scaffolds in genome assembly	577	176	241
G + C content (%)	51.49	51.49	49.91
Protein-coding genes	11415	10582	9849
Average protein length (aa)	486	506	493
Gene density (genes per Mbp)	295.7	271.1	258.8
Average exons per gene	2.8	2.8	2.7
Predicted secreted proteins	1413	1865	1490
tRNA genes	132	141	122
Repetitive DNA (%) - <i>de novo</i> repeats	2.4	2.12	4.42
Repetitive DNA (%) - RepBase (taxon: "fungi")	0.98	1.05	1.21
Overall RIP dominance (CpA $\leftrightarrow$ TpA)	1.0	0.22	1.14
Overall TA/AT (repetitive DNA)	1.02	1.02	1.31
Overall TA/AT (non-repetitive DNA)	0.7	0.69	0.77

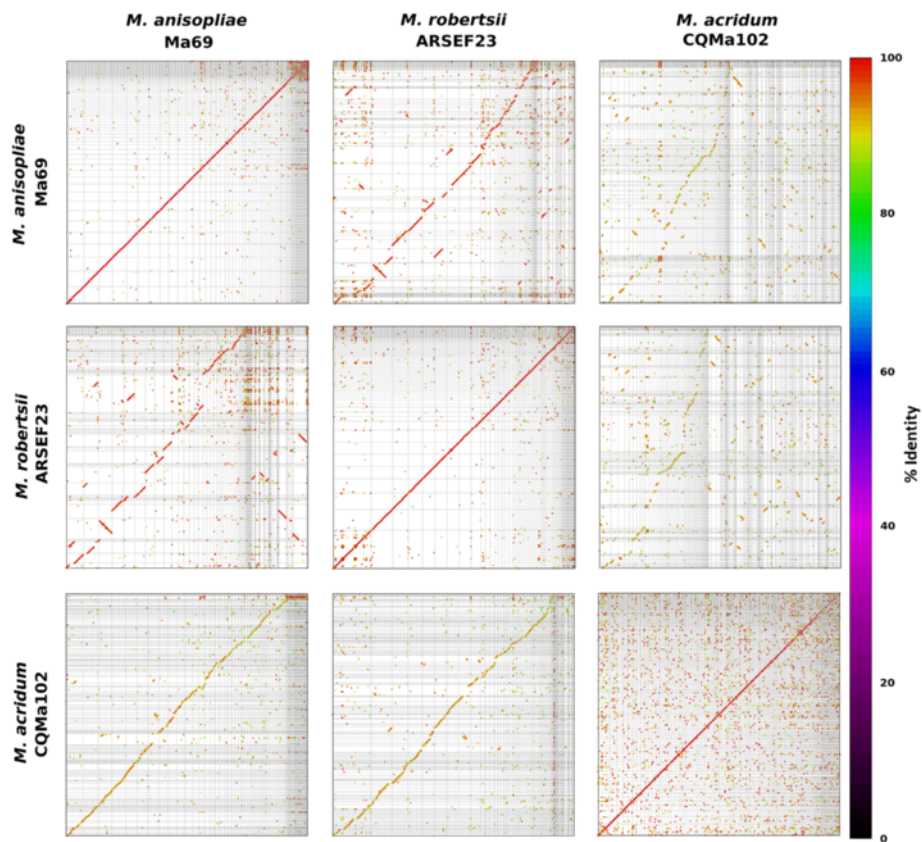
whole genome shotgun project of the organism *M. anisopliae*, isolate BRIP 53293 EFD69 SSC31 (Ma69), was deposited at DDBJ/EMBL-Bank/GenBank under the accession APNB000000000.

### Whole-genome synteny comparisons between *Metarhizium* species

Synteny dot-plots were generated for pair-wise comparisons between the genomes of Ma69, ARSEF23 and CQMa102. Significant macrosynteny and high levels of sequence homology ( $\geq 95\%$  sequence identity) were observed between the genomes of *M. anisopliae* and *M. robertsii* (Figure 3). Macrosynteny was also observed between *M. anisopliae* and *M. acridum* as well as between *M. robertsii* and *M. acridum*, albeit with lower levels of sequence homology (at  $\sim 90-95\%$  sequence identity).

### Gene prediction and predicted protein orthology

A total of 11,415 protein-encoding genes were predicted within the *M. anisopliae* Ma69 genome assembly,



**Figure 3** Whole-genome dot-plot comparison between *M. anisopliae* (Ma69), *M. robertsii* (ARSEF23) and *M. acridum* (CQMa102). The genome assemblies of all three genomes were compared via MUMMER 3.23 (nucmer). Regions of homology are plotted as diagonal lines or dots which are colour coded for percent identity (colour bar, right). Optimal co-linear order of scaffold sequences was determined by mummerplot (parameter: --fat, MUMMER 3.23) and differs in each pairwise comparison, however match coordinate data for all comparisons is available in Additional file 18.

compared with 10,582 and 9,849 from *M. robertsii* and *M. acridum*, respectively [13]. Orthology relationships between the translated proteins of the three *Metarhizium* species were predicted. While the number of predicted proteins in *M. anisopliae* was greater than those of *M. robertsii* and *M. acridum* (an increase of 833 and 1,566 respectively), most of these appeared to belong to gene families that are expanded in *M. anisopliae* (Additional file 2). A total of 127 proteins from *M. anisopliae* were predicted to have no orthologs in either *M. robertsii* or *M. acridum* (Additional file 3). These proteins were 'unique-by-orthology' to *M. anisopliae* and are referred to in this study as 'divergent'. Additionally, groups of orthologs in which the number of proteins belonging to one species was greater than the other two were classified as 'expanded'. There were 297 expanded groups containing 603 proteins in *M. anisopliae*, 257 groups containing 562 proteins in *M. robertsii* and 250 groups containing 540 genes in *M. acridum* (Additional file 4).

#### Functional annotation of species-specific (divergent) genes, expanded gene families and effector-like proteins in *M. anisopliae*, *M. robertsii* and *M. acridum*

Orthology relationships between the predicted proteins of the three *Metarhizium* species were used as a basis for predicting 'divergent' and 'expanded' gene families. Expanded gene families in all 3 species were abundant in proteins with similar Pfam annotations. These were generally related to membrane transport, sugar metabolism, protein kinases, cytochrome p450s, and transcription factors (Additional file 5). Putative functional annotations were assigned to divergent genes by comparison to multiple databases and algorithms, including: BLASTp versus NCBI Protein and Swissprot, gene ontologies (GOs), Interpro, Pfam, SignalP, WolfPsort and BioPerl::SeqStats. Functional annotations were then 'manually curated' based on the sum collection of supporting evidence for each gene, with a view, where possible, to intelligibly describe its putative role in pathogenicity. For the purposes of summarizing this analysis, the divergent

genes were then sorted into generalized categories based on their putative role and/or function. Of the 127 genes found to be divergent in *M. anisopliae*, 56 (44.1%) contained motifs known to be associated with pathogenicity in other species, 8 (6.3%) had homology to genes in PHI-base [15], 70 (55.1%) were low-molecular weight proteins of  $\leq 30$  kDa and 10 (7.87%) had predicted signal peptides (Additional file 6). The divergent genes of *M. anisopliae* were found, compared to its complete gene content, to be relatively more abundant in genes encoding membrane-anchored proteins, transposable elements, or having unknown functions and other functions described in more detail in sub-sections below. Genes encoding DNA/RNA binding factors and degradative enzymes were more abundant in the divergent genes of *M. robertsii* relative to its complete gene content (Additional file 7).

#### DNA/RNA binding – transcriptional regulation

Thirteen divergent genes of *M. anisopliae* were broadly characterized as having DNA/RNA binding functions, that is, potentially involved in the regulation of gene expression and intracellular signaling. No signal peptides were identified on these genes. Four of these genes were putative helicases, 2 were putative PIWI-like argonaut/dicer proteins, 4 had endonuclease/exonuclease/phosphatase (EEP) domains, 2 were putative fungal transcriptional factors and one was a putative ribonuclease H homolog.

#### Degradative enzymes

*M. anisopliae* had six divergent genes that were grouped in three pairs of paralogs that encoded enzymes with a degradative function. The products of gene paralogs *Ma69\_03389/Ma69\_04038* were putative peptidases (GO: 0008238). Gene paralogs *Ma69\_06536/Ma69\_11051* also encoded putative peptidases, each with a signal peptide and with homology to a *Phytophthora sojae* GIP1-like effector in PHIbase (PHI:653). Gene paralogs *Ma69\_00354/Ma69\_02206* encoded putative subtilisins. *Ma69\_00354* in particular was assigned GO terms indicating: serine type endopeptidase (GO: 0004252); active evasion of host immune response via regulation of host complement system (GO: 0042874); alkaline serine protease alp1 (GO: 0005576) and pathogenesis (GO: 0009405).

#### Membrane-anchored proteins

Twelve genes, divergent in *M. anisopliae* were classified as encoding membrane-anchored proteins. These genes included a putative arrestin-like G-protein coupled receptor, several membrane-associated proteins and a major facilitator superfamily (MFS) membrane transporter protein. Two genes were homologous to CaNAG4, a membrane transport protein (PHI: 511) and CaMDR1, a multidrug transporter protein (PHI: 26) both found in *Candida albicans*.

#### Transposable elements, other genes and unknown genes

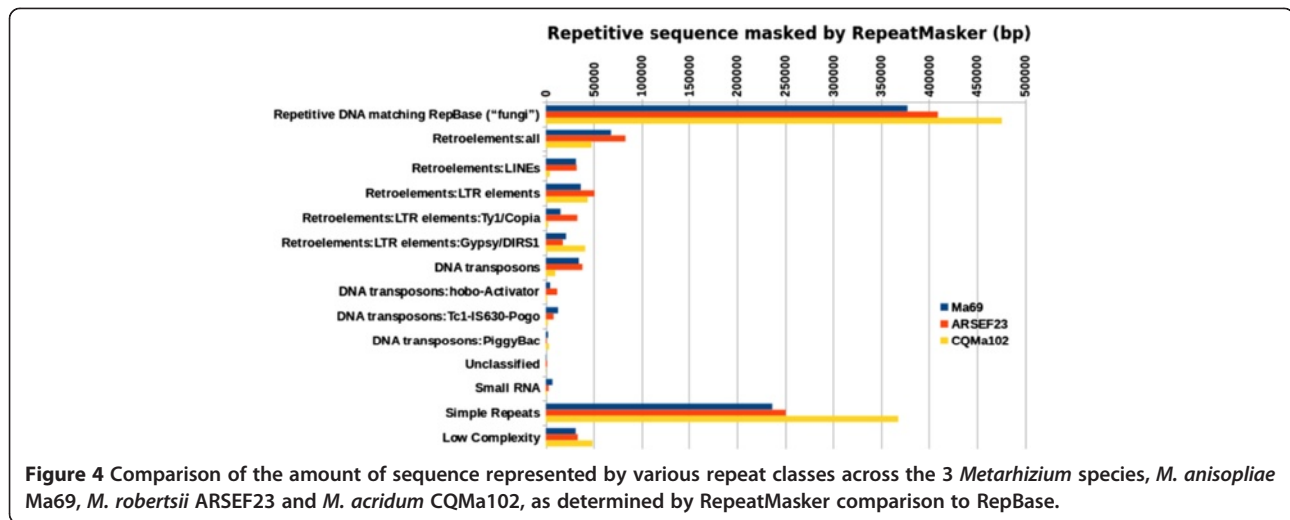
Fourteen genes divergent in *M. anisopliae* were identified as transposons and were therefore not considered of interest to this study. Of the remaining divergent genes in *M. anisopliae*, 44 had functions including protein kinases, histidine kinases and heat-shock proteins. Four divergent genes had homology to fungal pathogenicity genes in PHIbase. Two of these were homologous to TOXF from *Cochliobolus carbonum* and two had homology to CTB3 from *Cercospora nicotianae* (PHI: 157 and PHI: 1051 respectively). Due to a lack of homology to the datasets screened, 38 divergent genes, lacked functional annotations and as such their putative biological roles are unknown. Two of these unknown genes had predicted signal peptides.

#### Divergent candidate secreted effector proteins in *M. anisopliae* Ma69

Candidate secreted effector proteins are defined here as small protein molecules ( $\leq 300$  amino acids in length) which are putatively capable of being secreted by the pathogen via the eukaryotic secretory pathway. Secretion to the apoplast relies on the presence of an N-terminal signal peptide and some of these molecules are transferred into plant cells, facilitated by an amino acid motif, downstream of the signal peptide. Of the 127 divergent genes in *M. anisopliae*, 10 were found to encode putative signal peptides. Nine of these were  $\leq 300$  amino acids in length, four had pathogenicity motifs and these proteins were deemed candidate secreted effector proteins. These candidate secreted effector proteins were assigned to various functional categories including: degradative enzymes (3), membrane anchored proteins (2), other (2) and unknown genes (2).

#### Repetitive DNA content and repeat-induced point mutation (RIP)

Repetitive DNA analysis was performed on the scaffold sequences of *M. anisopliae* and the published reference genomes of *M. robertsii* and *M. acridum*. Repetitive sequences were predicted *de novo* and RIPCAL was used to determine genome-wide dinucleotide frequencies and to quantify RIP-like polymorphisms (SNP mutation biased towards CpA  $\rightarrow$  TpA or its reverse complement TpG  $\rightarrow$  TpA) within alignments of each repeat family (using the deRIP consensus as the model sequence for comparison) [16,17]. To facilitate comparison of repeat types between *Metarhizium* spp., their respective genomic matches to characterized repeat sequences in RepBase [18] are also shown in Figure 4. Overall, the repetitive contents of the genome of *M. anisopliae* Ma69 and *M. robertsii* ARSEF23 were similar and relatively low in comparison to other Pezizomycotina. In contrast, the *M. acridum* genome assembly contained approximately twice as much repetitive



DNA. In total, 2.4% of the genome assembly was predicted to be repetitive for *M. anisopliae*, 2.12% for *M. robertsii*, and 4.42% in *M. acridum*. In the three *Metarhizium* genomes, the most abundant repeat type was simple/low-complexity repeats, followed by retroelements and DNA transposons. The increased repetitive content of *M. acridum* was largely due to increased amounts of simple/low-complexity sequences. Interestingly, *M. acridum* was also relatively depleted in transposons and though maintaining a comparable level of LTR retroelements was relatively deficient in LINEs and DNA transposons.

Repeat families were scanned for repeat-induced point mutation (RIP)-like dinucleotide changes using two methods. The first involved genome-wide comparison of dinucleotide frequencies between *de novo*-identified repetitive sequences and non-repetitive sequences (Additional file 8). The RIP index TpA/APT, which measures the frequency of the common RIP-product TpA versus a non-RIP like control ApT, was 1.02 within repeats of *M. anisopliae* and *M. robertsii* and 1.31 in *M. acridum*. Dinucleotide frequency analysis also showed an overall increase in TpA dinucleotides in *M. acridum* relative to *M. anisopliae* and *M. robertsii* and a corresponding depletion of the RIP-target dinucleotides CpA and TpG (Additional file 8).

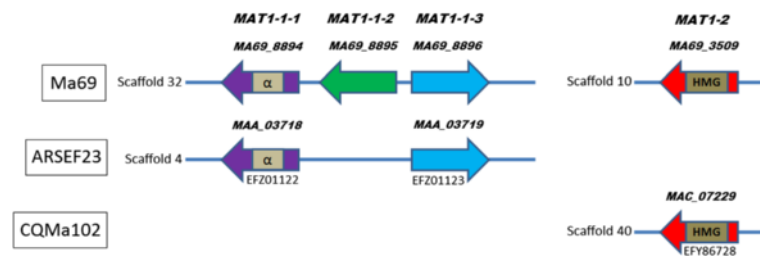
The second RIP-quantitation method used the RIP-CAL alignment-based method [16] versus a 'deRIPped' consensus of each family as a reference for comparison [17]. RIP mutation statistics for all repeat families in all three species are tabled in Additional file 9. RIP levels were on the whole relatively low in all three species, however all three species had a *rid* (cytosine-5 methyltransferase) homolog [19]. RIPCAL analysis showed that *M. anisopliae* and *M. robertsii* had similar total levels of RIP mutations and both species exhibited elevated levels of mutation of CpA dinucleotides (as well as CpT in some cases) typical of RIP in the Pezizomycotina in some repeat

families. It should be noted that RIP mutation requires a minimum repeat length and we would not expect it to occur in repeats less than 400 bp in length [20,21]. However assemblies containing short scaffolds such as those used in this study means these *Metarhizium* assemblies are likely to contain a high number of short repeat families which are incomplete versions of full length repeats. A total of 40 repeat families of *M. anisopliae*, ranging in average repeat family length of 47 to 375 bp, had CpA↔TpA (RIP-like) dominance scores of 1 or greater (Additional file 9 - summary). In *M. robertsii*, 12 repeat families had a similar RIP-like dominance score, with lengths ranging from 54 to 173 bp. In contrast, *M. acridum* had 56 repeat families with a RIP-like dominance score ≥1, lengths ranging from 45 to 1225 bp. Interestingly, the total number of RIP-like mutations in *M. robertsii* (7624) (relative to 'deRIPped' repeat consensus sequences), was approximately five times fewer than those of *M. anisopliae* and *M. acridum* (34960 and 39934 respectively).

#### Mating (MAT) type gene analysis

Analysis of orthologous relationships between the genes of the three *Metarhizium* species in this study identified incongruence in the number and presence of MAT genes (Figure 5). Ma69 had three genes putatively identified as *MAT 1-1-1* (MA69\_8894), *MAT1-1-2* (MA69\_8895) and *MAT1-1-3* (MA69\_8896) with no orthologous pairs in *M. acridum*, and orthologs of *MAT 1-1-1* (MAA\_03718) and *MAT1-1-3* (MAA\_03719) in *M. robertsii*. Ma69 also had one gene (MA69\_3509) putatively identified as *MAT1-2* which had an ortholog (MAC\_07229) in *M. acridum*, but not in *M. robertsii*.

The four putative *MAT* genes from Ma69 were subjected to BLASTp analysis and were found to have near identical sequence homology to ascomycetous mating



**Figure 5** Presence and absence of mating type genes in *M. anisopliae* Ma69, *M. robertsii* ARSEF23 and *M. acridum* CQMa102.

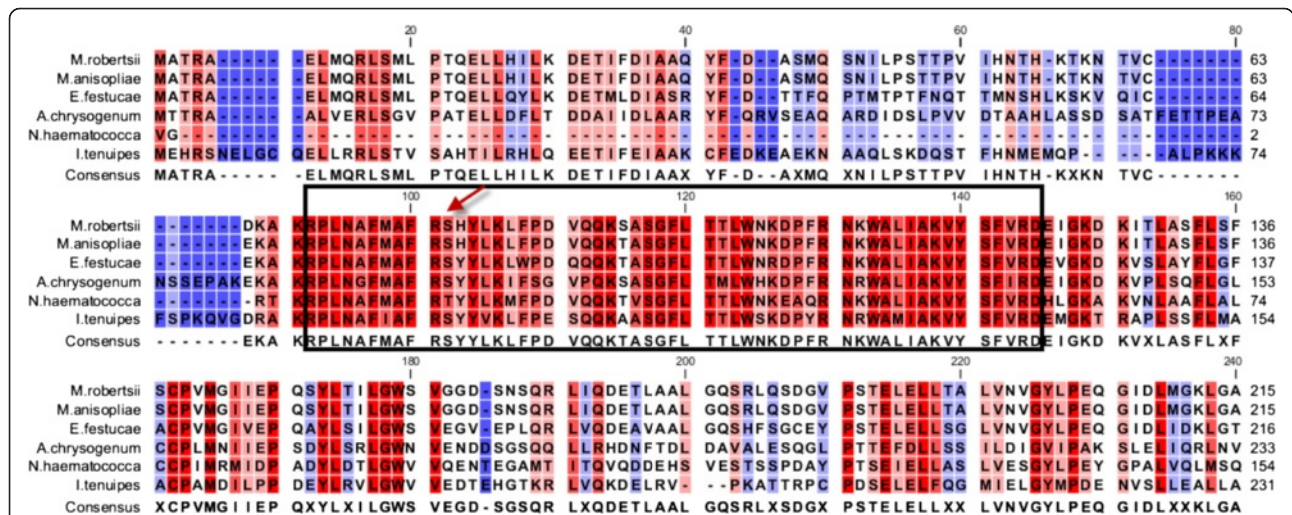
type (*MAT*) genes (Additional file 10). The top ten sequence hits for each *MAT* gene were aligned to confirm the identity of the genes and identify conserved domains (Additional file 11). The *M. anisopliae* *MAT1-1* alpha box domain [Pfam: PF04769] was confirmed by alignment to known *MAT1-1* sequences and carries a conserved intron (Figure 6) [22,23]. The *M. anisopliae* *MAT1-2* HMG-box domain [Pfam: PF00505] was also confirmed by alignment with other *MAT1-2* sequences and also carries a conserved intron (Figure 7) [22,24]. The presence of both *MAT* idiomorphs in the genome of the *M. anisopliae* isolate sequenced indicates it to be putatively homothallic.

**Prediction of secreted effector protein candidates via species-specificity and secretion predictions**

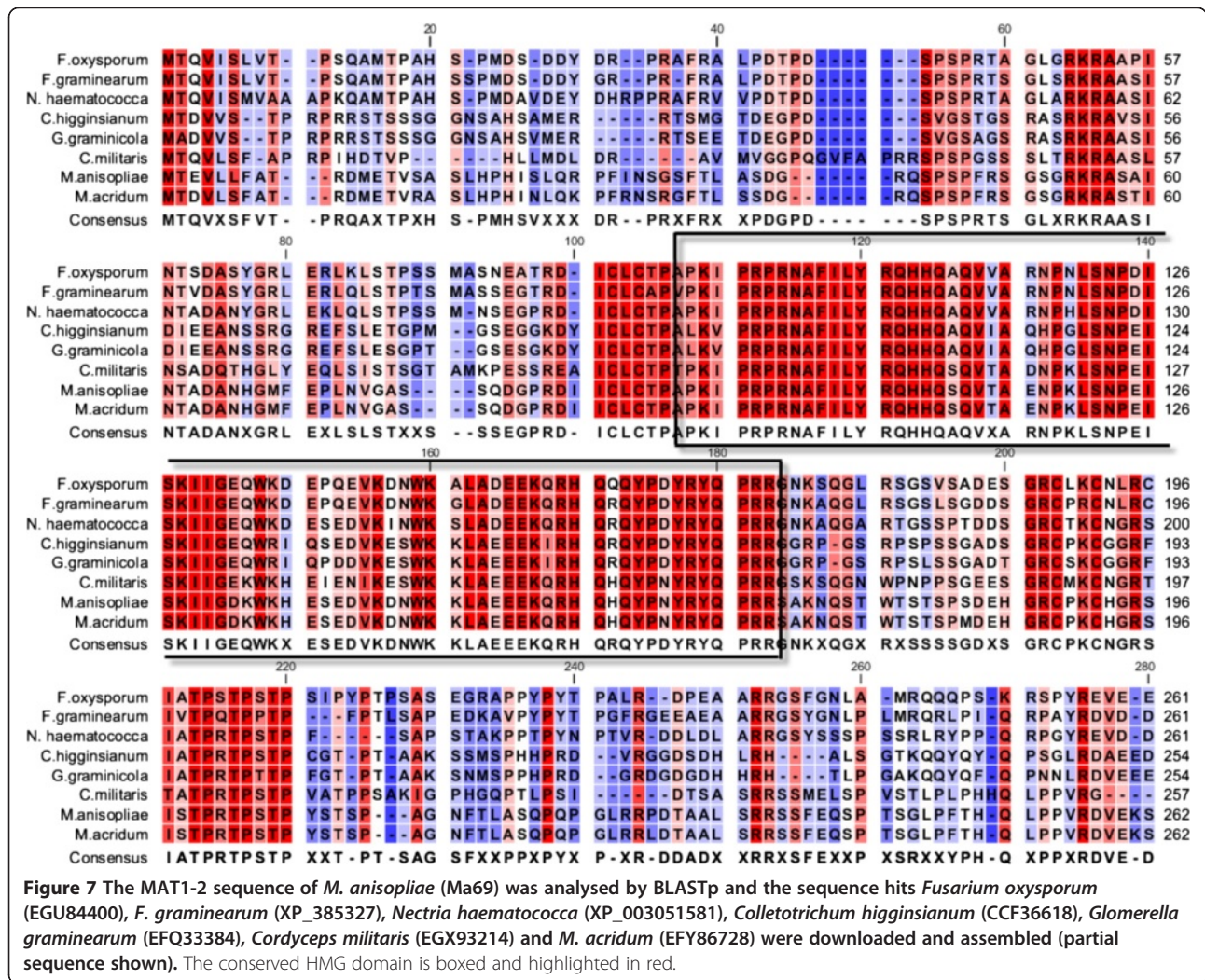
To our knowledge, pathogenicity effector-like motifs have not yet been examined in *Metarhizium* species. Some experimentally-validated pathogenicity effector proteins exhibit short conserved amino-acid motifs (Additional file 12), that in some cases have been demonstrated to

facilitate host-cell import [25]. We combined protein matches to these known ‘effector motifs,’ with predictions of signal-peptides (which infer extracellular secretion of the protein), to arrive at a conservative set of predicted secreted effector protein candidates.

In total, 1,620 proteins in *M. anisopliae* Ma69 contained sequences matching to a motifs associated with pathogenicity in other fungal species (Additional file 13). The most abundant motifs found were [YWF]XC, [LI]XAR and RXLR. These motifs are short (generally 3–4 amino acids), thus a substantial number of false-positive matches to proteins would be expected, in particular in long, intracellular proteins unrelated to pathogenicity. However these motifs were only considered and found in high numbers among groups of genes that had been previously bioinformatically-filtered for protein properties relevant to pathogenicity. Potential pathogenicity genes were identified using orthology as a basis for predicting species-specific (‘divergent’) genes in all 3 species, which are strong candidates as determinants of their respective differences in host-range phenotype. Seven amino acid



**Figure 6** The *MAT1-1-1* sequence of *M. anisopliae* (Ma69) was analysed by BLASTp and the sequences of *M. robertsii* (EFZ01122), *Epichloe festucae* (ACN59937), *Acremonium chrysogenum* (CAQ42990), *Nectria haematococca* (XP\_003052790) and *Isaria tenuipes* (BAC67541) were downloaded and aligned (partial sequence shown here). The conserved alpha box is boxed and highlighted in red and an arrow designates the conserved intron.



motifs known to be associated with pathogenicity proteins were found to match several divergent genes of *M. anisopliae*, *M. robertsii* and *M. acridum* (Additional file 14).

Furthermore, 1,295 proteins were predicted to be secreted in *M. anisopliae* Ma69 (Additional file 15). We identified 242 of these proteins that we term ‘candidate effectors’, which both contained a pathogenicity effector motif and were predicted to be secreted (Additional file 16). Among these 242 candidate effectors, 6 motifs were found, in descending order of abundance: [YFW]XC (166 proteins), [LI]XAR (54), RXLR (18), CHXC (2), KECXD (1) and YXSL[RK] (1).

## Discussion

*M. anisopliae* is a fungus with significant commercial and industrial applications as a biopesticide. The host and environmental range of *M. anisopliae* is regarded as entomologically cosmopolitan, compared to that of *M. acridum* which is host-specific. Recent molecular analysis of the *M.*

*anisopliae* complex has proposed its division into nine taxa, or distinct species, which individually may have narrower host specificities [5,8]. In a multigene phylogeny, *M. anisopliae* clustered with three other species *M. pingshaense*, *M. robertsii* and *M. brunneum* forming a distinct clade (PARB) [5]. Whether or not these represent discrete biological species or merely *formae specialis* is not known, as *M. anisopliae* has no known teleomorph [6] and the basis of its assumed asexuality has not been determined. Furthermore, advancing the currently poor understanding of the underlying mechanisms of genetic exchange could have important implications for understanding the persistence of traits of biopesticides in commercial applications.

Given that the isolate ARSEF23 sequenced as *M. anisopliae* has subsequently been re-classified as *M. robertsii*, and is sufficiently divergent from *M. anisopliae* [5,9], an alternative genome reference is required for this important biocontrol fungus to differentiate it from and complement the existing *Metarhizium* genomic resources. Thus,



we sequenced the genome of *M. anisopliae* isolate Ma69 using Illumina short-read sequence data, producing an annotated, *de novo* draft genome assembly of 577 scaffolds with an N50 of 11 and an N50 length of 1.24 Mb. The whole genome assembly and its predicted gene content was compared to the published genomes of *M. robertsii* (ARSEF23) and *M. acridum* (CQMa102) (Table 1), with particular focus on the species-specific genes of *M. anisopliae* Ma69, mating type and bioinformatic prediction of its candidate secreted effector proteins.

#### Whole genome synteny, homology and divergent genes in *M. anisopliae*

The comparative genomics of *M. robertsii* and *M. acridum* has been comprehensively examined [13] and we have refrained from duplicating previous efforts, except to include and compare *M. anisopliae* in that context. The previously published genome analysis of *M. robertsii* [13] showed syntenic conservation with gene clusters of *M. acridum* but did not present any whole-genome scale synteny comparisons. This was likely due to the poorer contiguity (as indicated by N50, Table 1) of the *M. acridum* genome assembly, which would prevent whole-genome scale synteny from being accurately observed. In this study, whole genome comparisons including the new assembly of *M. anisopliae*, in which the largest scaffolds are of a comparable level of contiguity with those of *M. robertsii*, indicate a generally macrosyntenic conservation pattern (Figure 3, Additional file 17 and Additional file 18). A low level of intra-chromosomal rearrangement is also observed, however this is similar to levels of degraded macrosynteny between other Pezizomycotina species of the same genus, such as previously observed between the Aspergilli [26]. We found that both *M. anisopliae* and *M. robertsii* had nucleotide sequence identities  $\geq 90\%$  to *M. acridum*, but sequence identity to each other was  $\geq 95\%$ , confirming that *M. anisopliae* and *M. robertsii* are more closely related to each other than to *M. acridum*.

A total of 11,415 proteins were predicted in the *M. anisopliae* genome assembly, 833 and 1,566 more than *M. robertsii* and *M. acridum* respectively. We examined the orthology of genes within the three *Metarhizium* species and found 127 genes in Ma69 that had no orthologs in *M. robertsii* and *M. acridum*, which we refer to in this study as 'divergent' genes. We also defined a set of 'expanded' genes in which the number of inparalogs of a single species was greater than corresponding outparalogs in the other two species, of which for *M. anisopliae* there were 603 expanded genes within 297 ortholog groups.

#### Repetitive DNA, TE classes and RIP analysis

Repetitive sequences are associated with transposable elements (TEs) which play a central role in the evolutionary restructuring of fungal genomes due to their ability

to move within the host genome, causing a range of mutations [27]. Excluding deleterious insertions, the mutational activity of TEs may promote genetic diversity and speed up adaptive evolution in the host [28]. Class I (retroelements) use a 'copy and paste' mechanism to transpose via the reverse-transcription of an RNA intermediate and include long terminal repeats (LTRs), non LTRs, and long and short interspersed nuclear elements (LINEs and SINEs, respectively) [27,29]. Class II TEs (DNA transposons) 'cut and paste' directly through a DNA form, using the enzyme transposase [27,29]. Earlier, we detailed our observations of phenotypic variation within single spored cultures of *M. anisopliae* (Figure 2). Due to their ability to directly excise from DNA and re-insert elsewhere in the genome, DNA transposons can generate a wide range of DNA sequence variation which may result in phenotypic changes and may account for the variety of cultural phenotypes observed. Alternatively, the abundance of Class II TEs in *M. anisopliae* may also affect cultural morphology, as evidence by LINE mediated mutations causing changes to conidial pattern formation in *Magnaporthe grisea* [30].

At the gene level, insertion of TEs either in or adjacent to genes may cause partial or total gene inactivation, resulting in new phenotypes [27,29]. On a genome-wide scale, TEs may be associated with large scale chromosomal modifications such as deletions, inversions and translocations [27]. To prevent or minimize potential deleterious effects of TEs, some fungi possess a gene silencing mechanism known as repeat induced point (RIP) mutation which targets duplicated DNA sequences  $>400$  bp long with  $>80\%$  shared identity [31]. Comparisons of the repeat content of the three *Metarhizium* spp., using reciprocal blast clustering, revealed surprisingly few common repetitive sequences. Further analysis of RIP-like polymorphism in all three *Metarhizium* spp. confirmed a common dinucleotide bias for CpA which conforms with expectations for species of the Pezizomycotina. As found in a number of species of Pezizomycotina [32], all three *Metarhizium* species examined here possessed a *rid* homolog which is essential for RIP and although the function of these genes is untested, we assume that RIP is active in all three species. In silico analysis of repetitive DNA and RIP across the three *Metarhizium* spp. indicated that repetitive content in *M. anisopliae* and *M. robertsii* was low at around 2% of genomic DNA and doubled in *M. acridum* at around 4%. Dinucleotide frequency analysis supports active RIP in all 3 species, with elevated frequencies of the RIP-product TpA and depleted levels of the primary RIP targets CpA and TpG within their respective repetitive DNA complements. *M. acridum* also exhibited a significant increase in TpA and decrease in CpA and TpG relative to the other two species, perhaps indicative of its higher transposon content. However a more complex picture emerges after alignment-based RIPCAL analysis, which supports similar levels of

RIP mutations between *M. anisopliae* and *M. acridum*, with a predicted 5-fold decrease in RIP-like polymorphism in *M. robertsii*. We speculate that while *M. anisopliae* and *M. robertsii* appear to contain similar levels of repetitive DNA, RIP activity may be greater in *M. anisopliae* than in *M. robertsii*. Higher numbers of RIP-like polymorphism in *M. anisopliae* and *M. acridum* may suggest a slightly greater adaptive potential compared to *M. robertsii*, as pathogenicity-related genes have previously been demonstrated to be affected by RIP mutations leaking outwards from flanking repetitive DNA [33,34].

### MAT gene orthology

Sexual function in filamentous ascomycetes is determined by mating type loci (*MAT*) which has been extensively described [35-37]. Typically, the single locus determining mating behaviour between mating partners contains different genes which are not allelic and are therefore known as idiomorphs [35,36]. In filamentous ascomycetes (Pezizomycotina), *mat* genes encode DNA binding motifs (high-mobility group (HMG) boxes and  $\alpha$  domains) [24] and are responsible for the control of both mating and incompatibility, cell-cell recognition and recognition between nuclei [35]. The lack of an observed sexual lifestyle in *Metarhizium* species may be the result of a loss of gene function, the lack of an opposite mating type, or merely the inability to induce a teleomorph under laboratory conditions. Functional *MAT* genes have been discovered in other ascomycetes previously assumed to be asexual (formerly known as Deuteromycetes) but the potential for a sexual cycle in these remains enigmatic [37-40] and indeed, its manifestation may occur on a continuum of sexuality in fungi, ranging from common to rare [41].

We identified both *MAT1-1* and *MAT1-2* idiomorphs in *M. anisopliae*, indicating that it is putatively homothallic and possibly capable of sexual reproduction. The Ma69 *MAT1-1* idiomorph encodes three proteins: *MAT1-1-1*, an  $\alpha$  domain protein; *MAT1-1-2*, an amphipathic  $\alpha$ -helical protein; and *MAT1-1-3*, an HMG box protein [42]. At this point, the functionality of these genes in *M. anisopliae* is not known, however sequence analysis has yielded a high level of conservation of alpha box and HMG domains with other ascomycetes. This warrants further investigation, including gene expression and/or transformation of closely related teleomorph species.

The *M. acridum* isolate CQMa102 possessed the *MAT1-2* gene but lacked the *MAT1-1* idiomorph. The *MAT1-2* gene encoded an HMG-domain protein that was highly conserved with *M. anisopliae* and other ascomycetes. Until an opposite mating type can be found and functionality is confirmed, heterothallism in *M. acridum* is deemed putative at this juncture. In contrast to both *M. anisopliae* and *M. acridum*, the *M. robertsii* isolate ARSEF23 possessed an incomplete *MAT1-1* idiomorph and lacked an ortholog to

*MAT1-2*. Absence of the *MAT1-2* idiomorph would indicate that the isolate sequenced was *MAT1-1* heterothallic, however again, in the absence of an opposite mating type and without confirmation of functionality, heterothallism is deemed putative. The relatively similar extents of RIP-like polymorphism in *M. anisopliae* and *M. acridum*, given the putative homothallism of *M. anisopliae*, is perplexing. RIP occurs only during meiosis, and as such, evidence of RIP may indicate the occurrence of meiosis in an ancestral or cryptic sexual stage, or alternatively, the existence of a similar process in vegetative cells [27]. The results of RIP analysis for *M. robertsii*, which indicate approximately 5 times less RIP than in the other *Metarhizium* spp., lends support to impaired or less frequent sexual activity in this species.

The missing *MAT1-1-2* gene in *M. robertsii*, a phenomenon also observed in *Pyrenopeziza brassicae* and *Cochliobolus heterostrophus*, two heterothallic species with sexual stages [37] is intriguing. A functional *MAT1-1-1* gene is critical to mating identity, sexual development [42] and vegetative incompatibility [43] as the alpha box domain, which is subsequently processed into mature pheromone molecules is located here [44]. However the requirement for a functional *MAT1-1-2* is less clear. In the aforementioned example of *P. brassicae*, the absence of *MAT1-1-2* does not impede out crossing, indeed the teleomorph is found naturally in oilseed rape [45]. In another example, *Neurospora crassa*, the homologs of *MAT1-1-2* and *MAT1-1-3* (reported as *matA-2* and *matA-3*, respectively) were reported to be non-essential for mating or ascospore production [43,46], however their expression increased the efficiency of sexual development [43]. It was also reported that homologs of *MAT1-1-2* and *MAT1-1-3* in *Podospora anserina* (reported as  $\alpha$ -helical genes and HMG-1 genes, respectively) were not required for mating but were required for sexual development and biparental progeny, as mutations in the  $\alpha$ -helical gene lead to barren fruiting bodies [37,47]. The evidence suggests that while *MAT1-1-1* is crucial to mating, the *MAT1-1-2* gene may have a supporting role in some ascomycetes in the successful development of sexual bodies, post mating, while in other species, its absence does not impede the development of fit progeny. The effect of the absence of *MAT1-1-2* in *M. robertsii* is unknown at this point, however it is intriguing that its closely related sibling species, *M. anisopliae* possesses the full complement of *MAT1-1* genes. In future, the discovery of isolates of *M. robertsii* and *M. acridum* with complementary *MAT* idiomorphs followed by evolutionary analysis of these idiomorphs may help to further our understanding of the role of *MAT1-1-2* in the *Metarhizium* species complex, and the potential effect on genetic exchange and perhaps, the observable phenotypic variation between cultures.

### Candidate secreted effector proteins

In this analysis, we identified a suite of genes which have the characteristics of effector proteins. Effector proteins are defined as molecules produced by a pathogen, which can alter host cell structure or function, thereby facilitating infection and/or initiating defense mechanisms [48]. Putative effector proteins are characterized in this study as being  $\leq 300$  amino acids in length and have a predicted signal peptide, which would facilitate secretion into the pathogen's extracellular space. In addition to these characteristics, effector predictions may also be supported by matches to proteins with previously identified pathogenicity motifs.

Proportionately, the genome of *M. anisopliae* had similar numbers of putative secreted effector-like proteins to *M. robertsii*, and more secreted effector-like proteins than *M. acridum* suggesting a similar capability or role of secreted proteins between *M. anisopliae* and *M. robertsii*. Ten predicted secreted effectors in *M. anisopliae* were divergent from *M. robertsii* and *M. acridum*. In particular, 3 of these divergent secreted effectors were determined to be putative degradative enzymes, which is of significant biological interest to *Metarhizium* biopesticide research and warrants further investigation to identify their function and potential effect on host range.

We found six motifs known to be associated with pathogenicity in other species, to be present in 242 candidate secreted effector-like proteins of *M. anisopliae*. Of these, 166 matched [YFW]xC, 54 matched [LI]XAR and 18 matched RXLR and represent the first analysis of effector-associated motifs in an entomopathogenic fungus. The best characterized of these motifs is the RXLR motif found in the oomycete *Phytophthora infestans* [49], although conserved RXLR motif effectors have yet to be observed in fungi. In pathogenicity effectors of *P. infestans*, the RXLR motif is found adjacent to the signal peptide, on the N-terminal end of the mature cleaved protein, and has been shown to facilitate translocation of the effector protein across the host membrane into the host cell. The [LI]XAR motif was identified in effector proteins secreted by the rice blast pathogen *Magnaporthe oryzae*, however the function of this motif is unknown [50]. As such, the function of these motifs in *Metarhizium* remains speculative.

While the high number of [YFW]XC motifs among the two predicted sets of genes, divergent and effector candidates, may simply be an artifact of random background matches to this short, three residue motif, there is some evidence suggesting that there may be an important, albeit, still unknown role for [YFW]XC motifs in secreted proteins. The [YFW]XC motif was first reported in the powdery mildew fungus *Blumeria graminis* [51]. It was predominantly expressed in the haustoria and was also over-represented among its predicted

secretome [51]. Since then, the [YFW]XC motif has also been discovered among effector candidates from the haustoria-producing rust fungi *Puccinia graminis* f.sp. *tritici*, *P. striiformis* f.sp. *tritici* and *Melampsora larici-populina* (as cited by Pedersen). *M. anisopliae* also produces haustoria *in vitro* under nutrient deprivation [52] however the link between the [YFW]XC motif and production of haustoria is still yet to be resolved. The function of the [YFW]XC motif still remains unclear, however it may have arisen from an extracellular RNase ancestor [53] and may be involved in establishing disulphide bridges with a C-terminal cysteine residue (Thordal-Christensen pers. comm.) thereby assisting protein folding and enhancing extracellular stability. Pedersen et al., [53] hypothesize that a secreted fungal ribonuclease appears to be the common origin of many of their candidates for secreted effector proteins (CSEPs). They suggested that some of these CSEPs could still be involved in interactions with host RNAs and modulate host immunity via this route. They also go on to suggest that extracellular ribonucleases are very stable molecules, resistant to proteolytic degradation, thereby providing a rigid scaffold ideal for evolving an effector arsenal, in which exposed loop regions subjected to positive selection allow diversification and evasion of host recognition. The structural conservation among effector candidates from diverse plant pathogenic species supports the hypothesis of an ancient common ancestor. The entomopathogenic *M. anisopliae* is likely to have arisen from a plant-pathogenic predecessor [13] and the discovery of these motifs in *Metarhizium* is consistent with the high level of conservation and putative functional requirement of these for pathogenicity.

## Conclusions

### Genomics, phenotypic variability and biopesticides

The key to successful commercialization of biopesticides is the identification and selection of virulent microbial isolates which are amenable to industrial scale production and formulation [54]. Information from comparative genomics studies can be used to identify genes which may contribute to isolate virulence and fitness as well as other characteristics which affect cultural variability and stability of potential biopesticides.

The discovery of both MAT idiomorphs in *M. anisopliae* raises more questions about its perceived asexuality, the potential pathways of genetic exchange in this species, impact on virulence, and implications for industrial applications. Understanding the molecular basis of mating genes not only gives insight into fundamental processes such as evolution of homothallism, heterothallism and asexuality, but also facilitates research on ascomycetous species of industrial interest and subsequent applied aspects [55]. Proof of function or not in these genes, will enhance the understanding of genetic exchange pathways in this species.

Overall, the repetitive contents of the genome of *M. anisopliae* Ma69 and *M. robertsii* ARSEF23 were similar and relatively low in comparison to other Pezizomycotina. In contrast, the *M. acridum* genome assembly contained approximately twice as much repetitive DNA. RIP levels were on the whole relatively low in all three species, however all three species had a *rid* (cytosine-5 methyltransferase) homolog and we observed SNP polymorphism consistent with active RIP in all three species. *M. anisopliae* and *M. robertsii* had similar total levels of RIP-like mutations and both species exhibited elevated levels of mutation of CpA dinucleotides (as well as CpT in some cases) typical of RIP in the Pezizomycotina in some repeat families. Interestingly, the total number of RIP-like mutations in *M. robertsii* was approximately five times fewer than those of *M. anisopliae* and *M. acridum*, lending support to impaired or less frequent sexual activity in this species. RIP occurs only during meiosis, and as such, evidence of RIP in these species may indicate the occurrence of meiosis in an ancestral or cryptic sexual stage, or alternatively, the existence of a similar process in vegetative cells.

The availability of an annotated whole genome sequence for *M. anisopliae* adds value to the already published genomes of *M. robertsii* and *M. acridum*; however in and of itself, represents a significant resource for future research into this agriculturally important fungal biopesticide. The nomenclature of the species *M. robertsii* needs to be widely adopted immediately to prevent further confusion with *M. anisopliae*, therefore the publication of the genome reference of *M. anisopliae* will serve as a valuable reference to differentiate it from *M. robertsii*.

## Methods

### Origin and culture of fungal strain Ma69

An Australian isolate of *M. anisopliae* known to be pathogenic to aphids was selected for whole genome sequencing. Originally, *M. anisopliae* isolate BRIP 53293 was isolated from soil in the Kingaroy region, Queensland and was obtained from the Queensland Department of Employment, Economic Development and Innovation (DEEDI). A single spore culture of BRIP 53293 (BRIP 53293 EFD 69 SSC31) was prepared on Sabouraud Dextrose Agar (SDA) before being transferred to Sabouraud Dextrose Broth (SDB), where cultures were shaken (150 rpm) at 25°C for 5 days prior to DNA extraction. For brevity, isolate BRIP53293 EFD69 SSC31 will be referred to hereafter as Ma69.

### Genome sequencing and assembly

Genomic DNA was extracted from the single-spored fungal culture using NucleoSpin Plant II (Machery Nagel) DNA extraction kit as per the manufacturer's instruction. The ITS region of ribosomal DNA (rDNA) was sequenced [56] and subjected to BLAST [57] analysis to confirm the

identity of the isolates. Sufficient genomic DNA was then prepared for paired-end and mate-pair sequencing at the Australian Genome Research Facility (AGRF), Brisbane according to Illumina protocols. Raw read sequences were trimmed via cutadapt [58] for: Illumina adaptor sequences; PCR primer and barcode contaminant sequences; homopolymers and runs of unknown bases > 5 bp; base quality > Q30 and; a minimum read length after trimming  $\geq$  50 bp. Reads with a discarded pair (i.e. with a length of < 50 bp after trimming) were removed from paired-end datasets, but were retained as singleton reads.

### De novo assembly of Ma69

Paired-end reads were assembled *de novo* with Velvet (version 1.1.2) [59]. The k-mer length parameter was tested between 20 and 70 bp in 2 bp increments and optimized for minimal N50 and maximal N50 length. Velvet paired-end scaffolds were then scaffolded with reads from the 3 kb mate-paired library via SSPACE (version 1.1) [60].

### Comparative genomics: whole assembly alignment

Whole genome alignments between the genome assemblies of Ma69, ARSEF23 and CQMa102 were performed with MUMMER 3.23 [61], using the nucmer algorithm with the `-mum` parameter. Dot-plots were generated with mummerplot with the `-filter`, `-colour` and `-fat` parameters.

### Gene prediction

Putative protein-coding regions of the Ma69 assembly were predicted with GeneMark-ES [62]. Predicted gene models were filtered for a minimum protein product length of 50 amino acids. Predicted genes encoding products < 50 amino acids in length were discarded.

### Protein family classification and orthology

Predicted proteins of Ma69 were compared for orthology to the ARSEF23 and CQMa102 published protein datasets, obtained from the NCBI NR protein database. Orthologous relationships between genes of these three species were inferred with ProteinOrtho4 using the `-pair` and `-selfblast` parameters [63]. A list of divergent genes specific to each species was generated according to non-orthology. In order to assign putative functions to the species-specific gene datasets, the datasets were screened and compared to a number of databases and were then manually-curated based on the sum collection of evidence for each gene. The genes were then sorted into generalized categories based on their putative role/function. The whole proteome datasets for each species were screened via EMBOSS for sixteen motifs known to be involved in fungal pathogenicity (Additional file 9). Homology to PHIBase (v 3.2) protein sequences were assessed by BLASTp (e-value threshold  $1e-5$ ). The best BLAST hit was assigned on the basis of highest bit score

and the disruption phenotype was recorded. Gene Ontology (GO) terms were assigned to the species-specific gene translated sequences via Blast2GO (default settings with BLASTp, Annex augmentation). Interproscan was also enabled within Blast2GO assigning additional GO term, functional domain and structure annotations. Secretion, sub-cellular localization, molecular weight and iso-electric point were calculated for each species-specific dataset via SignalP 3.0 [64], WolfPsort [65] and EMBOSS IEP [66]. Amino acid compositions were calculated using the SeqStats BioPerl module [67].

### Repetitive DNA analysis

*De novo* prediction of repetitive sequences of Ma69 and the published reference genomes of ARSEF23 (*M. robertsii*) and CQMa102 (*M. acridum*) were generated using RepeatScout [68] while overlapping and redundant consensus repeats were combined via CAP3 [69]. These *de novo* repeats were mapped to the genome assembly via RepeatMasker [70]. For the purpose of comparison of characterised repeat types, RepeatMasker was also run versus each *Metarhizium* spp. assembly using RepBase sequences corresponding to the “fungi” taxon. Multiple alignments of repeat families were generated by RIPCAL [16] using ClustalW [71,72]. deRIP [17] was used to predict the pre RIP consensus of each aligned repeat family. RIPCAL [16] was run on each repeat family alignment using the deRIP consensus as a model sequence for comparison. Repeats for all species were analyzed using TEClass [73] to predict the most probable class for each repeat family. Repeats common between all species were identified using proteinortho4 using the ‘-p blastn’ parameter [63].

### Mating (MAT) type gene analysis

MAT genes were identified by homology to characterized mating-type sequences using BLASTP [57]. The amino-acid sequences top ten BLASTp hits for each putative Ma69 MAT protein were obtained from NCBI and multiple-alignment on these was performed in CLC bio Genomics Workbench v4.0 (CLC bio, Denmark).

### Search for known effector-associated motifs

Proteins were assessed for the presence of domains known to be associated with effectors outlined in Additional file 9 using EMBOSSpreg v 6.5.7 (<http://emboss.sourceforge.net/apps/release/6.3/emboss/apps/preg.html>).

### Additional files

**Additional file 1:** Top ten BLAST hits for Ma69 (complete sequence ITS1; 5.8S rRNA gene; ITS2).

**Additional file 2:** *Metarhizium* ortholog table: Orthology relationships between the translated proteins of the three

*Metarhizium* species were predicted with ProteinOrtho4 using the -pair and -selfblast parameters.

**Additional file 3:** *Metarhizium* species-specific gene summary: A total of 127 proteins from *M. anisopliae* were predicted to have no orthologs in either *M. robertsii* or *M. acridum*. These proteins were ‘unique-by-orthology’ to *M. anisopliae* and are referred to in this study as ‘divergent’.

**Additional file 4:** Orthologs and expanded genes: Groups of orthologs in which the number of proteins belonging to one species was greater than the other two were classified as ‘expanded’. There were 297 expanded groups containing 603 proteins in *M. anisopliae*, 257 groups containing 562 proteins in *M. robertsii* and 250 groups containing 540 genes in *M. acridum*.

**Additional file 5:** Pfam functional annotations assigned to *Metarhizium* proteins.

**Additional file 6:** *Metarhizium* species-specific gene summary: Putative function annotations were assigned to divergent genes by comparison to multiple databases and algorithms, including: BLASTp versus NCBI Protein and Swissprot, gene ontologies (GOs), Interpro, Pfam, SignalP, WolfPsort and BioPerl::SeqStats. Functional annotations were then ‘manually curated’ based on the sum collection of supporting evidence for each gene, with a view to intelligibly describe its putative role in pathogenicity. For the purposes of summarizing this analysis, the divergent genes were then sorted into generalized categories based on their putative role and/or function.

**Additional file 7:** Functional classification of divergent gene in *M. anisopliae*, *M. robertsii* and *M. acridum*: Gene annotations listed for which amino acid translations (predicted proteins) were found by reciprocal-best hit analysis via Proteinortho between the three species.

**Additional file 8:** Summary of RIP dinucleotide analysis of *de novo* identified repetitive sequences and non-repetitive sequences.

**Additional file 9:** RIP mutation statistics for all repeat families in all three species: Repeat families were scanned for repeat-induced point mutation (RIP)-like dinucleotide changes using two methods. The second RIP-quantitation method used the RIPCAL alignment-based method versus a ‘deRIPped’ consensus of each family as a reference for comparison.

**Additional file 10:** BLASTp identification of putative MAT genes from *M. anisopliae* isolate Ma69.

**Additional file 11:** Top 10 blastn sequence hits for Ma69 MAT genes.

**Additional file 12:** A list of 16 known amino-acid motifs involved in plant-pathogenicity in other species were used to match the whole proteome datasets of *M. anisopliae*, *M. robertsii* and *M. acridum* via EMBOSS (Preg).

**Additional file 13:** 1620 Ma69 proteins found with motifs linked with pathogenicity in other species.

**Additional file 14:** Seven known motifs linked with pathogenicity in other species were matched to the three *Metarhizium* species investigated.

**Additional file 15:** 1295 putative proteins in Ma69 found with signal peptides.

**Additional file 16:** 242 candidate effector secreted proteins found in Ma69.

**Additional file 17:** Whole genome synteny coordinates from *M. anisopliae*, *M. robertsii*, and *M. acridum*.

**Additional file 18:** Whole genome synteny figure comparing the distribution of scaffold length from all three *Metarhizium* assemblies.

### Competing interests

The authors declare that they have no competing interests.

#### Authors' contributions

JP, BW, BS and GA conceived of the study and participated in its design and coordination. JP carried out the laboratory work. JH and AW performed the bioinformatics analysis. All authors helped to draft and approved the final manuscript.

#### Acknowledgements

JP thanks Mr Daryn Webster for advice and assistance with spreadsheet manipulation. This work was funded by EH Graham Centre for Agricultural Innovation, an alliance between Charles Sturt University and New South Wales Department of Primary Industries. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

#### Author details

<sup>1</sup>Graham Centre for Agricultural Innovation, School of Agricultural and Wine Sciences, Charles Sturt University, Locked Bag 588, Wagga Wagga 2650, NSW, Australia. <sup>2</sup>Black Box Bioinformatics, Perth, WA, Australia. <sup>3</sup>Centre for Crop and Disease Management, Department of Environment and Agriculture, Curtin University, Perth, WA, Australia.

Received: 17 December 2013 Accepted: 17 July 2014

Published: 7 August 2014

#### References

- Shan LT, Feng MG: Evaluation of the biocontrol potential of various *Metarhizium* isolates against green peach aphid *Myzus persicae* (Homoptera: Aphididae). *Pest Manag Sci* 2010, **66**:669–675.
- Roberts DW, St Leger RJ: *Metarhizium* spp., Cosmopolitan Insect-Pathogenic Fungi: Mycological Aspects. In *Advances in Applied Microbiology*. Academic Press; 2004:1–70. doi:10.1016/S0065-2164(04)54001-7.
- Vestergaard S, Gillespie AT, Butt TM, Schreiter G, Eilenberg J: Pathogenicity of the hyphomycete fungi *Verticillium-lecanii* and *Metarhizium-anisopliae* to the western flower thrips, *Frankliniella-occidentalis*. *Biocontrol Sci Tech* 1995, **5**:185–192.
- Zimmermann G: Review on safety of the entomopathogenic fungus *Metarhizium anisopliae*. *Biocontrol Sci Tech* 2007, **17**:879–920.
- Kepler RM, Sung G-H, Ban S, Nakagiri A, Chen M-J, Huang B, Li Z, Spatafora JW: New teleomorph combinations in the entomopathogenic genus *Metacordyceps*. *Mycologia* 2012, **104**(1):182–197.
- Driver F, Milner RJ, Trueman JWH: A taxonomic revision of *Metarhizium* based on phylogenetic analysis of rDNA sequence data. *Mycol Res* 2000, **104**(2):134–150.
- Taylor JW, Jacobson DJ, Fisher MC: The evolution of asexual fungi: reproduction, speciation and classification. *Annu Rev Phytopathol* 1999, **37**:197–246.
- Bischoff JF, Rehner SA, Humber RA: A multilocus phylogeny of the *Metarhizium anisopliae* lineage. *Mycologia* 2009, **101**(4):512–530.
- Inglis GD, Duke GM, Kabaluk JT, Geotzel MA: Genetic diversity of *Metarhizium anisopliae* var. *anisopliae* in southwestern British Columbia. *J Invertebr Pathol* 2008, **98**:101–113.
- Wang CS, Butt TM, St. Leger RJ: Colony sectorization of *Metarhizium anisopliae* is a sign of ageing. *Microbiology - SGM* 2005, **151**:3223–3236.
- Ryan MJ, Bridge PD, Smith D, Jeffries P: Phenotypic degeneration occurs during sector formation in *Metarhizium anisopliae*. *J Appl Microbiol* 2002, **93**:163–168.
- Wang S, O'Brien TR, Pava-Ripoll M, St. Leger RJ: Local adaptation of an introduced transgenic insect fungal pathogen due to new beneficial mutations. *Proc Natl Acad Sci U S A* 2011, **108**(51):20449–20454.
- Gao Q, Jin K, Ying S-H, Zhang Y, Xiao G, Shang Y, Duan Z, Hu X, Xie X-Q, Zhou G, Peng G, Luo Z, Huang W, Wang B, Fang W, Wang S, Zhong Y, Ma L-J, St Leger RJ, Zhao G-P, Pei Y, Feng M-G, Xia Y, Wang C: Genome sequencing and comparative transcriptomics of the model entomopathogenic fungi *Metarhizium anisopliae* and *M. acridum*. *PLoS Genet* 2011, **7**(1):e1001264.
- Ryan MJ, Smith D, Bridge PD, Jeffries P: The relationship between fungal preservation method and secondary metabolite production in *Metarhizium anisopliae* and *Fusarium oxysporum*. *World J Microbiol Biotechnol* 2003, **19**:839–844.
- Winnenburg R, Urban M, Beacham A, Baldwin TK, Holland S, Lindeberg M, Hansen H, Rawlings C, Hammond-Kosack K, Kohler J: PHI-base update: additions to the pathogen-host interaction database. *Nucleic Acids Res* 2008, **36**(SI):D572–D576.
- Hane JK, Oliver RP: RIPCAL: a tool for alignment-based analysis of repeat-induced point mutations in fungal genomic sequences. *BMC Bioinformatics* 2008, **9**:478.
- Hane JK, Oliver RP: In silico reversal of repeat-induced point mutation (RIP) identifies the origins of repeat families and uncovers obscured duplicated genes. *BMC Genomics* 2010, **11**:655.
- Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J: Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* 2005, **110**:462–467.
- Freitag M, Williams RL, Kothe GO, Selker EU: A cytosine methyltransferase homologue is essential for repeat-induced point mutation in *Neurospora crassa*. *Proc Natl Acad Sci U S A* 2002, **99**(13):8802–8807.
- Watters MK, Randall TA, Margolin BS, Selker EU, Stadler DR: Action of repeat-induced point mutation on both strands of a duplex and on tandem duplications of various sizes in *Neurospora*. *Genetics* 1999, **153**(2):705–714.
- Cambareri E, Singer M, Selker E: Recurrence of repeat-induced point mutation (RIP) in *Neurospora crassa*. *Genetics* 1991, **127**:699–710.
- Woo PCY, Chong KTK, Tse H, Cai JJ, Lau CCY, Zhou AC, Lau SKP, Yuen K-Y: Genomic and experimental evidence for a potential sexual cycle in the pathogenic thermal dimorphic fungus *Penicillium marneffei*. *FEBS Lett* 2006, **580**:3409–3416.
- Paoletti M, Rydholm C, Schwier EU, Anderson MJ, Szakacs G, Lutzoni F, Debeauvais J-P, Latge J-P, Denning DW, Dyer PS: Evidence for sexuality in the opportunistic fungal pathogen *Aspergillus fumigatus*. *Curr Biol* 2005, **15**:1242–1248.
- Poggeler S: Genomic evidence for mating abilities in the asexual pathogen *Aspergillus fumigatus*. *Curr Genet* 2002, **42**:153–160.
- Kale SD, Tyler BM: Entry of oomycete and fungal effectors into plant and animal host cells. *Cell Microbiol* 2011, **13**(12):1839–1848.
- Hane JK, Rouxel T, Howlett BJ, Kema GHJ, Goodwin SB, Oliver RP: A novel mode of chromosomal evolution peculiar to filamentous Ascomycete fungi. *Genome Biol* 2011, **12**:R45.
- Daboussi M-J, Capy P: Transposable elements in filamentous fungi. *Annu Rev Microbiol* 2003, **57**:275–299.
- Santana MF, Silva JCF, Batista AD, Ribeiro LE, da Silva GF, de Araujo EF, de Queiroz MV: Abundance, distribution and potential impact of transposable elements in the genome of *Mycosphaerella fijiensis*. *BMC Genomics* 2012, **13**:720.
- Favaro LCL, de Araujo WL, de Azevedo JL, Paccola-Meirelles LD: The biology and potential for genetic research of transposable elements in filamentous fungi. *Genet Mol Biol* 2005, **28**(4):804–813.
- Nishimura M, Hayashi N, Jwa N-S, Lau GW, Hamer JE, Hasebe A: Insertion of the LINE retrotransposon MGL causes a condioophore pattern mutation in *Magnaporth grisea*. *Mol Plant Microbe Interact* 2000, **8**:892–894.
- Novikova OS, Fet V, Vlinov AG: Homology-dependent inactivation of LTR retrotransposons in *Aspergillus fumigatus* and *A. nidulans* genome. *Mol Biol* 2007, **41**:886–893.
- Clutterbuck AJ: Genomic evidence of repeat-induced point mutation (RIP) in filamentous ascomycetes. *Fungal Genet Biol* 2011, **48**:306–326.
- Fudal I, Ross S, Brun H, Besnard AL, Ermel M, Kuhn ML, Balesdent MH, Rouxel T: Repeat-induced point mutation (RIP) as an alternative mechanism of evolution toward virulence in *Leptosphaeria maculans*. *Mol Plant Microbe Interact* 2009, **22**(8):932–941.
- Van de Wouw AP, Cozijnsen AJ, Hane JK, Brunner PC, McDonald BA, Oliver RP, Howlett BJ: Evolution of linked avirulence effectors in *Leptosphaeria maculans* is affected by genomic environment and exposure to resistance genes in host plants. *PLoS Pathog* 2010, **6**(11):e1001180.
- Coppin E, Debuchy R, Arnaise S, Picard M: Mating types and sexual development in filamentous ascomycetes. *Microbiol Mol Biol Rev* 1997, **61**(4):411–428.
- Kronstad JW, Staben C: Mating type in filamentous fungi. *Annu Rev Genet* 1997, **31**(1):245–276.
- Shiu PKT, Glass NL: Cell and nuclear recognition mechanisms mediated by mating type in filamentous ascomycetes. *Curr Opin Microbiol* 2000, **3**:183–188.
- Arie T, Kaneko I, Yoshida T, Noguchi M, Nomura Y, Yamaguchi I: Mating-type genes from asexual phytopathogenic Ascomycetes *Fusarium oxysporum* and *Alternaria alternata*. *Mol Plant Microbe Interact* 2000, **13**(12):1330–1339.

39. Yun SH, Arie T, Kaneko I, Yoder OC, Turgeon BG: **Molecular organisation of mating type loci in heterothallic, homothallic, and asexual *Gibberella/Fusarium* species.** *Fungal Genet Biol* 2000, **31**:7–20.
40. Sharon A, Yamaguchi K, Christiansen S, Horwitz BA, Yoder OC, Turgeon BG: **An asexual fungus has the potential for sexual development.** *Mol Gen Genet* 1996, **251**:60–68.
41. Dyer PS, Paoletti M: **Reproduction in *Aspergillus fumigatus*: sexuality in a supposedly asexual species?** *Med Mycol* 2005, **43**:S7–S17.
42. Glass N, Grotelueschen J, Metzzenberg R: ***Neurospora crassa* A mating-type region.** *Proc Natl Acad Sci U S A* 1990, **87**:4912–4916.
43. Ferreira AV, An Z, Metzzenberg RL, Glass NL: **Characterization of mat A-2, mat A-3 and deltamataA mating-type mutants of *Neurospora crassa*.** *Genetics* 1998, **148**(3):1069–1079.
44. Jones SK Jr, Bennett RJ: **Fungal mating pheromones: Choreographing the dating game.** *Fungal Genet Biol* 2011, **48**:668–676.
45. Lacey ME, Rawlinson CJ, McCartney HA: **First record of the natural occurrence in England of the teleomorph of *Pyrenopeziza brassicae* on oilseed rape.** *Trans Br Mycol Soc* 1987, **89**(1):135–140.
46. Bobrowicz P, Pawlak R, Correa A, Bell-Pedersen D, Ebbole DJ: **The *Neurospora crassa* pheromone precursor genes are regulated by the mating type locus and the circadian clock.** *Mol Microbiol* 2002, **45**(3):795–804.
47. Arnise S, Debuchy R, Picard M: **What is a *bona fide* mating-type gene? Internuclear complementation of *mat* mutants in *Posospora anserina*.** *Mol Gen Genet* 1997, **256**(2):169–178.
48. Kamoun S: **A catalogue of the effector secretome of plant pathogenic oomycetes.** *Annu Rev Phytopathol* 2006, **44**:41–60.
49. Whisson SC, Boevink PC, Moleleki L, Avrova AO, Morales JG, Gilroy EM, Armstrong MR, Grouffaud S, van West P, Chapman S, Hein I, Toth IK, Pritchard L, Birch PRJ: **A translocation signal for delivery of oomycete effector proteins into host plant cells.** *Nature* 2007, **450**:115–119.
50. Yoshida K, Saitoh H, Fujisawa S, Kanzaki H, Matsumura H, Yoshida K, Tosa Y, Chuma I, Takano Y, Win J, Kamoun S, Terauchi R: **Association genetics reveals three novel avirulence genes from the rice blast fungal pathogen *Magnaporthe oryzae*.** *Plant Cell* 2009, **21**:1573–1591.
51. Godfrey D, Bohlenius H, Pedersen C, Zhang Z, Emmersen J, Thordal-Christensen H: **Powdery mildew fungal effector candidates share N-terminal Y/F/WxC motif.** *BMC Genomics* 2010, **11**(1):317.
52. St. Leger RJ, Staples RC, Roberts DW: **Cloning and regulatory analysis of starvation-stress gene, *ssgA*, encoding a hydrophobin-like protein from the entomopathogenic fungus, *Metarhizium anisopliae*.** *Gene* 1992, **120**:119–124.
53. Pedersen C, Ver Loren van Themaat E, McGuffin LJ, Abbott JC, Burgis TA, Barton G, Bindschedler LV, Lu X, Maekawa T, Wessling R, Cramer R, Thordal-Christensen H, Panstruga R, Spanu PD: **Structure and evolution of barley powdery mildew effector candidates.** *BMC Genomics* 2012, **13**:694.
54. Ash GJ: **The science, art and business of successful bioherbicides.** *Biol Control* 2010, **52**:230–240.
55. Poggeler S: **Mating-type genes for classical strain improvements of ascomycetes.** *Appl Microbiol Biotechnol* 2001, **56**:589–601.
56. White TJ, Bruns T, Lee S, Taylor J: **Amplification and Direct Sequencing of Fungal Ribosomal RNA Genes for Phylogenetics.** In *PCR Protocols: A Guide to Methods and Applications*. Edited by Innis MA, Gelfand DH, Sninsky JJ, White TJ. San Diego: Academic Press; 1990:315–322.
57. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**:403–410.
58. Martin M: **Cutadapt removes adapter sequences from high-throughput sequencing reads.** *EMBnet J* 2011, **17**(1):10–12. <http://journal.embnet.org/index.php/embnetjournal/article/view/200>.
59. Zerbino DR, Birney E: **Velvet: Algorithms for de novo short read assembly using de Bruijn graphs.** *Genome Res* 2010, **18**(5):821–829.
60. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W: **Scaffolding pre-assembled contigs using SSPACE.** *Bioinformatics* 2010, **27**(4):578–579.
61. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL: **Versatile and open software for comparing large genomes.** *Genome Biol* 2004, **5**(2):R12.
62. Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M: **Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training.** *Genome Res* 2008, **18**:1979–1990.
63. Lechner M, Findeib S, Steiner L, Marz M, Stadler PF, Prohaska SJ: **Proteinortho: detection of (Co-)orthologs in large-scale analysis.** *BMC Bioinformatics* 2011, **12**:124.
64. Petersen TN, Brunak S, von Heijne G, Nielsen H: **SignalP 4.0: discriminating signal peptides from transmembrane regions.** *Nat Methods* 2011, **8**:785–786.
65. Horton P, Park K-J, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K: **WoLF PSORT: protein localization predictor.** *Nucleic Acids Res* 2007, **35**(S):W585–W587.
66. Rice P, Longden I, Bleasby A: **EMBOSS: the European molecular biology open software suite.** *Trends Genet* 2000, **16**(6):276–277.
67. Stajich JE, Block D, Boulez K, Brenner SE, Chervitz SA, Dagdigian C, Fuellen G, Gilbert JG, Korf I, Lapp H, Lehvaslaiho H, Matsalla C, Mungall CJ, Osborne BI, Pocock MR, Schattner P, Senger M, Stein LD, Stupka E, Wildinson MD, Birney E: **The Bioperl toolkit: Perl modules for the life sciences.** *Genome Res* 2002, **12**(10):1611–1618.
68. Price AL, Jones NC, Pevzner PA: **De novo identification of repeat families in large genomes.** *Bioinformatics* 2005, **21**(Suppl 1):i351–i358.
69. Huang W, Madan A: **CAP3: a DNA sequence assembly program.** *Genome Res* 1999, **9**(9):868–877.
70. *RepeatMasker Open-3.0.* <http://www.repeatmasker.org>.
71. Chenna R, Sugarwarra H, Koike T, Lopez R, Gibson TJ, Thompson JD: **Multiple sequence alignment with the Clustal series of programs.** *Nucleic Acids Res* 2003, **31**(13):3497–3500.
72. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG: **Clustal W and Clustal X version 2.0.** *Bioinformatics* 2007, **23**(21):2947–2948.
73. Abrusan G, Grundmann N, DeMester L, Makalowski W: **TEclass- a tool for automated classification of unknown eukaryotic transposable elements.** *Bioinformatics (Oxford, England)* 2009, **25**(10):1329–1330.

doi:10.1186/1471-2164-15-660

**Cite this article as:** Pattemore et al.: The genome sequence of the biocontrol fungus *Metarhizium anisopliae* and comparative genomics of *Metarhizium* species. *BMC Genomics* 2014 **15**:660.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

