

RESEARCH ARTICLE

Open Access

Genomic characteristics and comparative genomics analysis of *Penicillium chrysogenum* KF-25

Qin Peng^{1†}, Yihui Yuan^{1†}, Meiyang Gao^{1*}, Xupeng Chen^{1,2}, Biao Liu¹, Pengming Liu¹, Yan Wu¹ and Dandan Wu¹

Abstract

Background: *Penicillium chrysogenum* has been used in producing penicillin and derived β -lactam antibiotics for many years. Although the genome of the mutant strain *P. chrysogenum* Wisconsin 54-1255 has already been sequenced, the versatility and genetic diversity of this species still needs to be intensively studied. In this study, the genome of the wild-type *P. chrysogenum* strain KF-25, which has high activity against *Ustilaginoidea virens*, was sequenced and characterized.

Results: The genome of KF-25 was about 29.9 Mb in size and contained 9,804 putative open reading frames (*orfs*). Thirteen genes were predicted to encode two-component system proteins, of which six were putatively involved in osmolarity adaption. There were 33 putative secondary metabolism pathways and numerous genes that were essential in metabolite biosynthesis. Several *P. chrysogenum* virus untranslated region sequences were found in the KF-25 genome, suggesting that there might be a relationship between the virus and *P. chrysogenum* in evolution. Comparative genome analysis showed that the genomes of KF-25 and Wisconsin 54-1255 were highly similar, except that KF-25 was 2.3 Mb smaller. Three hundred and fifty-five KF-25 specific genes were found and the biological functions of the proteins encoded by these genes were mainly unknown (232, representing 65%), except for some *orfs* encoding proteins with predicted functions in transport, metabolism, and signal transduction. Numerous KF-25-specific genes were found to be associated with the pathogenicity and virulence of the strains, which were identical to those of wild-type *P. chrysogenum* NRRL 1951.

Conclusion: Genome sequencing and comparative analysis are helpful in further understanding the biology, evolution, and environment adaption of *P. chrysogenum*, and provide a new tool for identifying further functional metabolites.

Keywords: *Penicillium chrysogenum*, Genome, Comparative genome

Background

The filamentous fungus *Penicillium chrysogenum* has been widely used for producing penicillin and derived β -lactam antibiotics for more than 80 years [1]. The discovery of penicillin has greatly improved human health and promoted the development of the medical industry. In addition to producing penicillin, *P. chrysogenum* has exhibited abilities in others areas, including bioleaching, biological remediation, promoting plant growth, and producing non-

β -lactam antibiotics and antifungal agents [2-6]. According to previous reports, several *P. chrysogenum* strains produce secreted proteins, such as PAF, PgAFP, and PgChP, which inhibit the growth of opportunistic zoopathogens, plant-pathogenic fungi, and toxigenic molds [7-9]. With their high stability, effective inhibitory activity, and broad inhibition spectra, these three proteins could be effective antifungal agents in medicine and agriculture [10,11].

In 2008, van den Berg et al. reported the first sequence of the *P. chrysogenum* genome and genes that were responsible for key steps in penicillin production were identified [12]. The genome not only led to a deeper understanding of penicillin synthesis, but also provided a new tool for identifying additional metabolites [13]. The

* Correspondence: mygao@wh.iov.cn

†Equal contributors

¹Key Laboratory of Agricultural and Environmental Microbiology, Wuhan Institute of Virology, Chinese Academy of Sciences, Wuhan 430071, China
Full list of author information is available at the end of the article

sequenced *P. chrysogenum* strain Wisconsin 54-1255 was a model laboratory strain that was derived from wild-type NRRL 1951, which was isolated from infected cantaloupe [14,15]. As a mutant strain used in the laboratory, Wisconsin 54-1255 might be some genetic variations, such as reduced PahA activity, encoded by *pahA*, in the catabolism of phenylacetic acid (the side chain precursor for the synthesis of benzylpenicillin) [16]. Moreover, different *P. chrysogenum* isolates maintain diverse genetic backgrounds [17,18], and studying the genome sequences of other strains will provide more information on the genetic diversity of *P. chrysogenum*. Therefore, sequencing the genome of a wild-type *P. chrysogenum* strain is necessary.

P. chrysogenum KF-25 is a wild-type strain isolated from a soil sample by our laboratory. It shows high-anti-fungal activity against *Ustilaginoidea virens*, which causes false smut disease of rice and corn in humid areas [19], in contrast to the Wisconsin 54-1255 strain, which did not exhibit anti-fungal activity. This suggested that there might be differences in the genetic backgrounds of the two strains. To provide more genetic information on *P. chrysogenum* to identify additional active substances and to determine the critical genes involved in the biosynthesis of the active substances, we sequenced and analyzed the genome of KF-25. Comparative genome analysis of strain KF-25 with Wisconsin 54-1255 and the wild-type strain NRRL 1951 revealed significant genetic variance. We also analyzed the functions and distribution of the genes encoding several important proteins, including transporters, non-ribosomal peptide synthase, and two-component regulatory systems (TCRSs).

Results and discussion

Strain features

The colony morphology and anti-fungal activity of strains KF-25 and Wisconsin 54-1255 were investigated. Following grown on potato-Sucrose-agar (PSA) plates for 5 days, flavescent water drops were observed on the surface of KF-25 colonies, but not on Wisconsin 54-1255 (Figure 1a,d). Strain KF-25 also produced more yellow pigment than Wisconsin 54-1255 (Figure 1b,e). The anti-fungal activities of the two strains against *U. virens* strain UV-1 were tested, and results showed that strain KF-25 had a strong inhibitory effect on UV-1 (Figure 1c), while no anti-fungal activity was observed for strain Wisconsin 54-1255 (Figure 1f). The fermentation broth of KF-25 and Wisconsin 54-1255 was analyzed by using HPLC-DAD and an additional peak was observed at time point 7.28 min in the HPLC chromatogram of KF-25 (Figure 1g,h). The component was collected from time point 7 to 8 min and the collected component showed a high activity against UV-1 (data not shown). As strain KF-25 is a wild isolate and Wisconsin 54-1255 is a mutant strain derived from NRRL 1951 [13], the different origins might cause the different physiological features.

Genome sequence and annotation of *P. chrysogenum* KF-25

General genome features

The genome of *P. chrysogenum* KF-25 was sequenced by a shotgun approach using Hiseq 2000 (Illumina, California, USA) with a read length of 2×100 bp. The 29.9 Mb genome was covered by 194 scaffolds and composed of 1,459

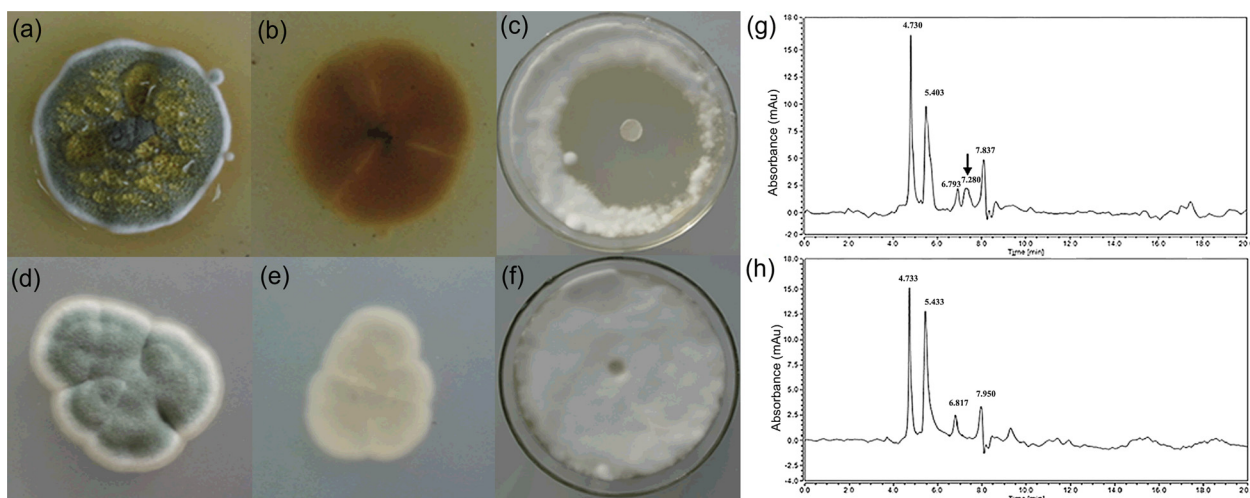


Figure 1 Colony features and anti-fungal activity of KF-25 and Wisconsin 54-1255. (a) and (d), the front colony morphology on PSA plates after 5 days of growth. (b) and (e), the back of the colony on PSA plates after 5 days of growth. (c) and (f), the inhibitory activity of the 4-day fermentation broths on *U. virens* strain UV-1. (g) and (h), the HPLC chromatogram of the four day fermentation broths. The upper row indicated the features of KF-25 and the lower row indicated the feature of Wisconsin 54-1255.

contigs with 154× coverage. Among the 194 scaffolds, the average length was 154 kb, with the largest being 2.72 Mb. The general features of the KF-25 genome compared with the Wisconsin 54-1255 genome are shown in Table 1. Genome annotation revealed that the genome of strain KF-25 encoded 9,804 ORFs, and that the GC content of the predicted protein-coding region was 53.4%. Among the 9,804 ORFs, 7044 were similar to proteins in UniProt database, 4,158 proteins were similar to proteins in the KEGG database (Figure 2), and 9,727 showed similarity to proteins from the NCBI nr database. Analysis of the 9,804 ORFs by KOGnitor indicated that 6,231 predicted proteins matched members of the eukaryotic orthologous groups (KOG) (Figure 2). In the genome of strain KF-25, 112 genes encoding tRNA and 29 rDNA genes were predicted using tRNAscan and RNAmmer software. The 112 tRNA genes were mainly scattered between scaffolds 1, 9, 15, 18, 19, 24 and 51, although sometimes four or five tRNA genes formed clusters. The anticodon usage of KF-25 is listed in Additional file 1: Table S1. Among the 9,804 predicted ORFs, 39 and 91 were identified as translation and transcription factors, respectively. (Additional file 1: Table S2 and Table S3).

In total, 317 repetitive elements were found in the genome of strain KF-25 by RepeatScout, with a minimum length 50 bp and a maximum length of 1,296 bp. Repetitive sequence analysis by using CENSOR indicated that 648,249 bp of the KF-25 genome (2.17%) was repeat sequences, while the repeat content of Wisconsin 54-1255 was 1.04% [12].

Microsatellites (simple sequence repeats, SSRs) are one of the most popular genetic markers and exist widely in fungal genomes. Because of high mutation rate and changing in repeat numbers during DNA replication, SSRs exhibit high individual specificity [20-22]. In the genome of KF-25, 3,798 SSRs were found, with sizes ranging from 15 to 167 bp, and these SSRs were homogeneously distributed throughout the genome (Additional file 1: Figure S1).

Table 1 General genome features of *P. chrysogenum* KF-25 and *P. chrysogenum* Wisconsin 54-1255

Genome features	<i>P. chrysogenum</i> KF-25	<i>P. chrysogenum</i> Wisconsin 54-1255 [12]
Assembly sizes (Mb)	29.9	32.2
GC content (%)	49.0	48.9
Gene number	9,804	12,943
Mean gene length (bp)	1,573	1,515
Exons per gene	3.2	3.0
Introns per gene	2.2	2.2
tRNA number	112	145
rRNA number	29	28

The secretory system and transporter

Translocation of protein and molecule across the plasma membrane is essential for cell life and requires the help of secretory systems and transporters, such as signal recognition particle (SRP) and the Sec translocase [23,24]. SRP plays a critical role in targeting of secretory proteins to the cellular membrane [25], while the Sec secretion system is responsible for protein translocation across the cytoplasmic membrane [26]. *P. chrysogenum* has been widely used to produce penicillin and some other secondary metabolites with antimicrobial activity [2,7-9,27,28]. The secretory system and transporters are essential for secretion of these antimicrobial substances and for import of their substrates. In the KF-25 genome, 12 proteins were predicted to be components of the eukaryotic Sec-SRP secretion systems (Additional file 1: Table S4). These proteins might play important roles in protein secretion in *P. chrysogenum*. Several genes in the genome of KF-25 encoded transporters or components of the secretion system that involved in producing penicillin and other secondary metabolites. KF-25 genome contained 531 genes that encoded transporter proteins, which mainly belonged to the major facilitator superfamily (MFS, 231 genes), and the ABC transporter superfamily (52 genes). Several genes in the secondary metabolism gene cluster were predicted to encode MFS-type transporters by antiSMASH [29]. The MFS transporters in the penicillin synthesis pathway could regulate the production of penicillin and enhance the sensitivity of *P. chrysogenum* to phenylacetic acid [30]. Many ABC superfamily transporters in the KF-25 genome were predicted to be multidrug resistance proteins [31]. One ABC superfamily transporter was reported to be critical in the export of phenylacetic acid, which is the precursor of penicillin synthesis. There were also several other transporters in the KF-25 genome that are involved in sugar, amino acid, cation, and vitamin transport.

Two-component regulatory system

TCRSs (Two-component regulatory systems) are found in bacteria, yeast, fungi and plant, and enable organisms to rapidly sense and adapt to specific environments [32]. TCRSs consisted of a sensor kinase and a response regulator, and are involved in regulating diverse processes, such as chemotaxis, osmolarity and differentiation [33-35]. According to previous reports, osmotic pressure regulates the morphogenesis and the secondary metabolism pathways of filamentous fungi via TCRSs [33,36,37]. Increased osmotic pressure stimulated the vegetative growth and conidia formation of *P. chrysogenum*, and also influenced its respiration and organic acid production [38,39]. The TCRSs that senses osmotic pressure and regulates the life cycle of *P. chrysogenum* might induce *P. chrysogenum* produce secondary metabolites, such as penicillin and other

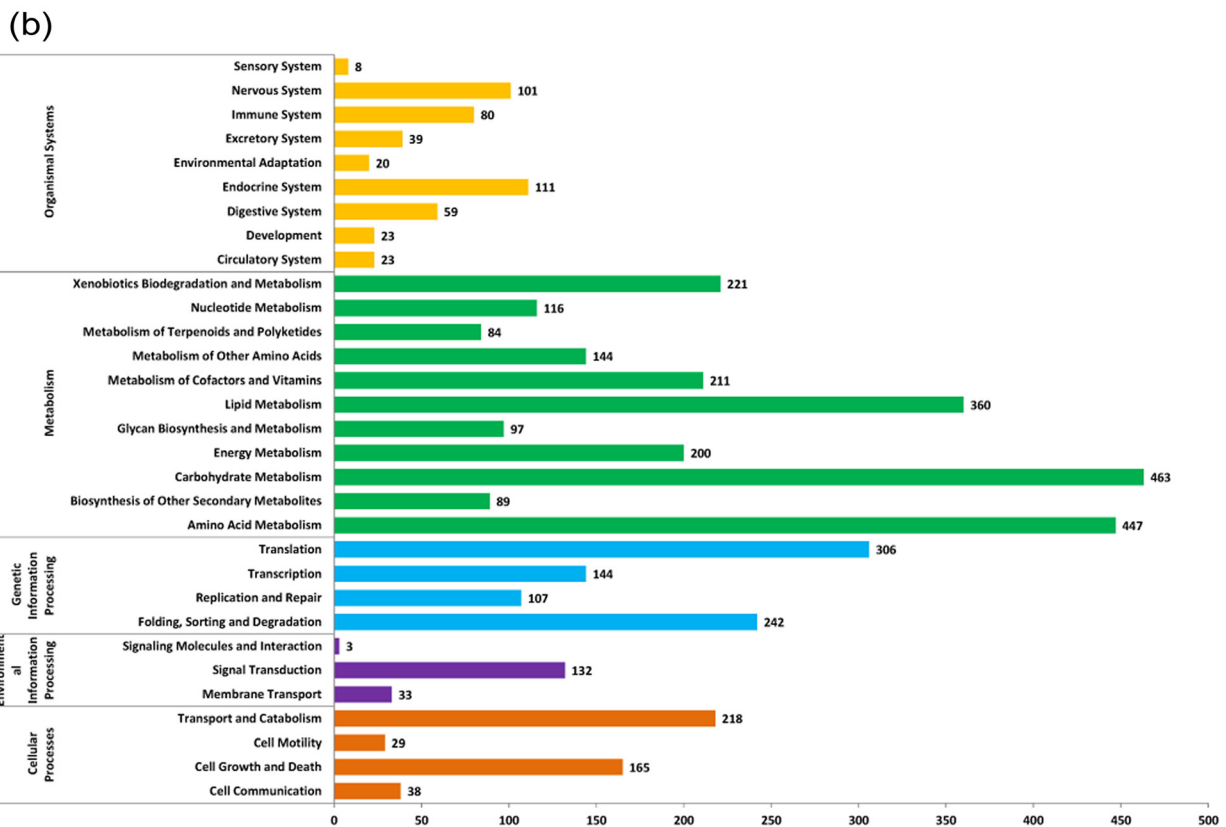
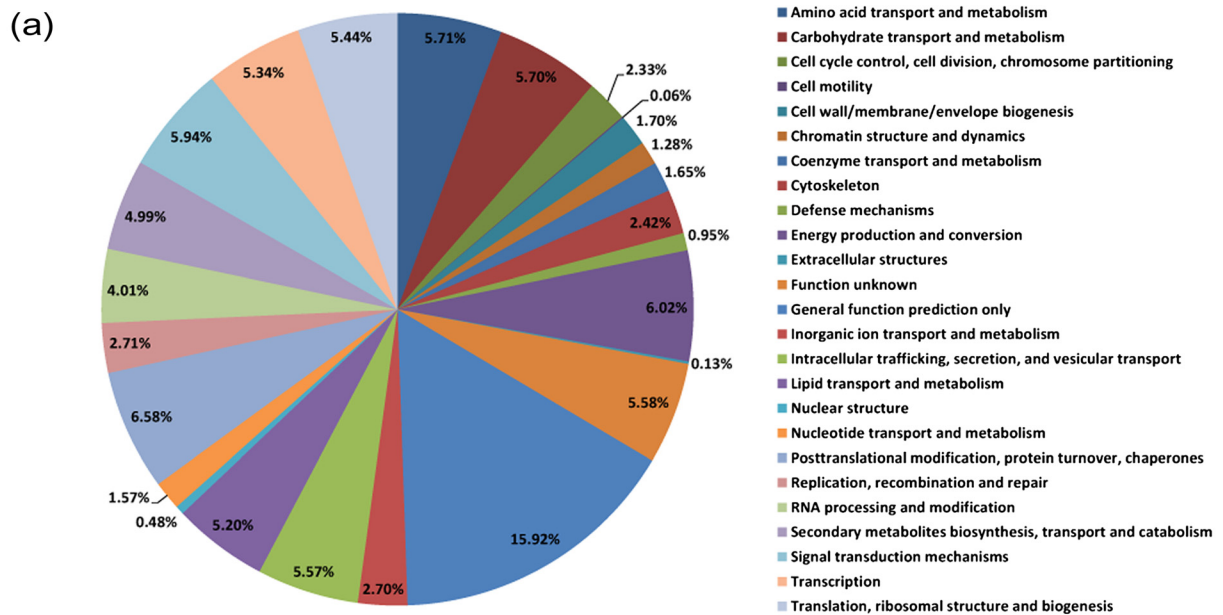


Figure 2 Characteristics of the genome of *P. chrysogenum* KF-25. (a) Functional classification of ORFs encoded by the genome of strain KF-25, based on the KOG (Eukaryotic Orthologous Groups of proteins) database. In total, 6,231 ORFs with orthologs in the KOG database were classified and the percentages indicate the frequencies of ORFs with assigned functions. **(b)** Functional classification of KF-25 genome ORFs based on the KEGG (Kyoto Encyclopedia of Genes and Genomes) database. In total, 4,158 ORFs had functional classifications assigned and the numbers with each classification are indicated.

bioactive agents. Thirteen predicted proteins based in KF-25 were involved in TCRSs. Among these proteins, four contained both the sensor kinase and response regulator domains, four contained only the sensor kinase domains, and the remaining five contained only the response regulator domains (Table 2). Six of the 13 predicted proteins were involved in TCRSs that sensed and adapted to the osmotic pressure. Similar proteins were also found in the genome of Wisconsin 54-1255 [12]. The existence of osmotic pressure-associated TCRSs in the *P. chrysogenum* genome might explain the ability of *P. chrysogenum* to adapt to high osmotic pressure. The other seven predicted TCRSs proteins were mainly involved in sensing or adapting to drugs, the cell cycle, or capsular synthesis.

Comparative genomics and phylogenetic analysis of *P. chrysogenum* KF-25

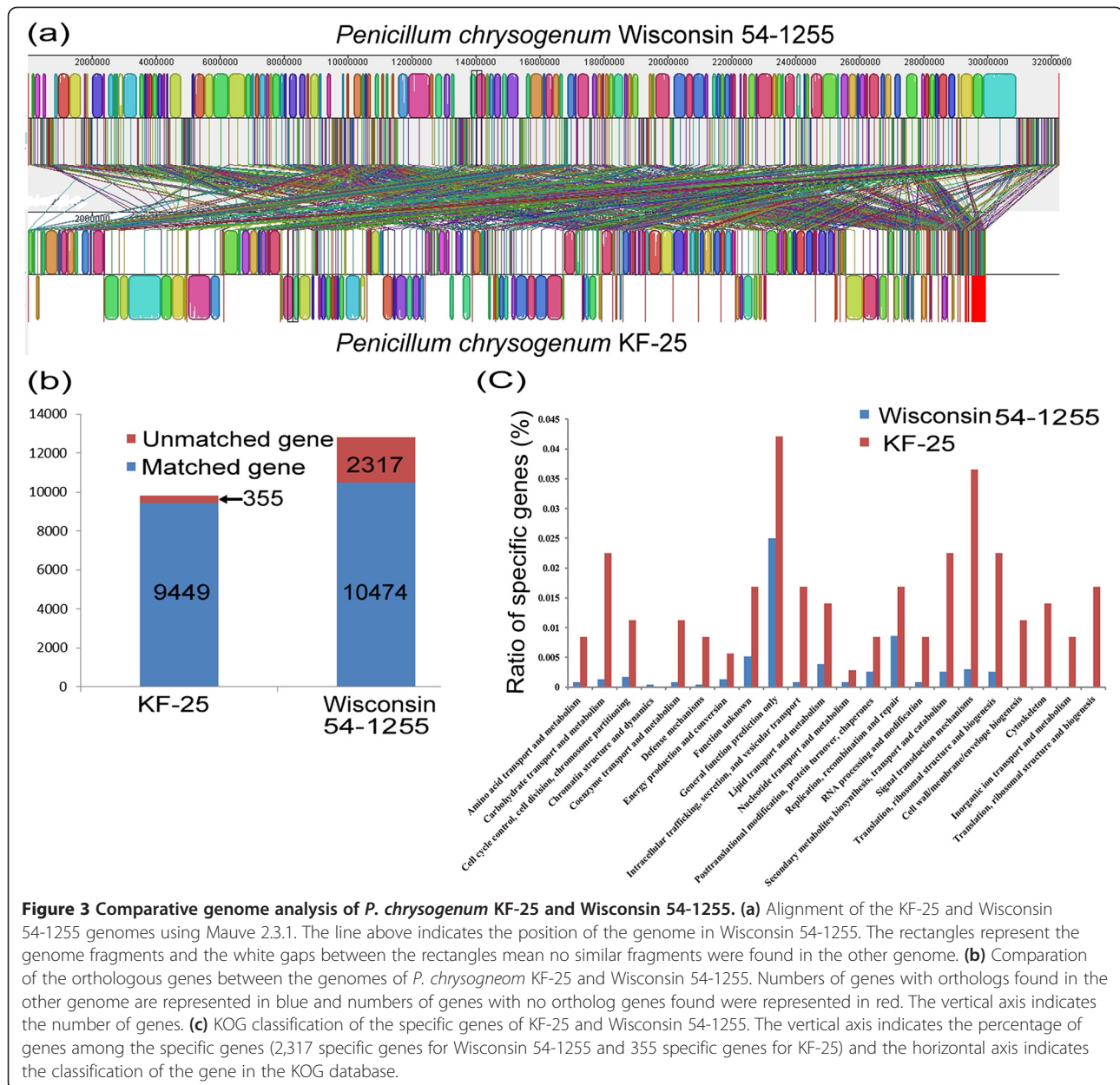
Comparative genome analysis of *P. chrysogenum* KF-25 and *P. chrysogenum* Wisconsin 54-1255

The genome of *P. chrysogenum* KF-25 was 2.3 Mb smaller than that of *P. chrysogenum* Wisconsin 54-1255 (Table 1). The genome of KF-25 was composed of 194 scaffolds, while the genome of Wisconsin 54-1255 was composed of only 49 super-contigs [12]. We speculated that gaps between the scaffolds might be one of the reasons for the smaller genome size of KF-25. Genomic alignment showed that the genome of KF-25 covered 93% of the Wisconsin 54-1255 genome. The average protein similarity between the predicted proteomes of KF-25 and Wisconsin 54-1255 was 75.1% (Figure 3a; Additional file 1: Figure S2). Several genome fragments,

with a total length of 2.3 Mb, were missing in the KF-25 genome. These fragments in the Wisconsin 54-1255 genome were mainly from the 5'-termini of contigs 13, 17, 23, 24 and the 3'-terminus of contig 22. According to a previous report, these fragments of the Wisconsin 54-1255 genome were not found in the genomes of other sequenced filamentous fungi, such as *Aspergillus nidulans*, *Aspergillus niger*, and *Aspergillus oxyaze* [12], and were proposed to contain the *P. chrysogenum*-specific genes [12]. Alignment of the proteomes of the two strains showed that 2,317 genes in the genome of Wisconsin 54-1255 were not found in the genome of KF-25 (Figure 3b), while 1,043 (representing 45%) of these genes were located in the 2.3 Mb of missing fragments. Based on these results, we inferred that these genes were not the *P. chrysogenum*-specific genes, but of Wisconsin 54-1255 strain-specific genes. The biological functions of most proteins encoded by these strain-specific genes are unknown (2183, representing 94.2%) [12], though some genes were involved in transport, metabolism, and transcription regulation (Figure 3c; Additional file 1: Table S5 and Figure S3A). The 2.3 Mb of missing fragments contained numerous repeat and transposable elements, and the introns in these regions were typically small and few compared with other regions of Wisconsin 54-1255 genome. Because the two sequenced *P. chrysogenum* strains were isolated from different geographical regions, and because Wisconsin 54-1255 is a laboratory strain that has undergone several rounds of mutation, the strain-specific sequences in the Wisconsin 54-1255 genome might have evolved by transposition and horizontal gene transfer.

Table 2 The genes encoding the proteins involving in the two-component systems in the genome of *P. chrysogenum* KF-25 and corresponding genes in the genome of *P. chrysogenum* Wisconsin 54-1255

Genes in Genome of KF-25	Putative protein function	Corresponding genes in <i>P. chrysogenum</i> Wisconsin 54-1255
KF25_0355	osmolarity two-component system, response regulator SSK1	Pc20g02430 (98% identify)
KF25_1660	two-component system, NarL family, capsular synthesis sensor histidine kinase RcsC	Pc22g18780 (99% identify)
KF25_2492	osmolarity two-component system, response regulator SKN7	Pc22g04440 (99% identify)
KF25_4368	two-component system, chemotaxis family, sensor kinase Cph1	Pc06g00040 (99% identify)
KF25_4886	two-component system, unclassified family, sensor histidine kinase and response regulator	Pc16g03520 (99% identify)
KF25_7115	two-component system, cell cycle sensor kinase and response regulator	Pc12g07950 (96% identify)
KF25_7139	two-component system, NarL family, capsular synthesis sensor histidine kinase RcsC	Pc22g07510 (99% identify)
KF25_7934	osmolarity two-component system, phosphorelay intermediate protein YPD1	Pc22g12510 (100% identify)
KF25_8216	osmolarity two-component system, response regulator SSK1	Pc22g16340 (99% identify)
KF25_8339	osmolarity two-component system, response regulator SSK1	Pc13g13580 (88% identify)
KF25_8360	osmolarity two-component system, response regulator SSK1	Pc13g13880 (99% identify)
KF25_9319	two-component system, unclassified family, sensor histidine kinase and response regulator	Pc13g09080 (99% identify)
KF25_9723	two-component system, unclassified family, sensor histidine kinase and response regulator	Pc20g15550 (99% identify)



Furthermore, there were 355 strain-specific genes in KF-25 genome that were not found in the genome of Wisconsin 54-1255 (Figure 3b). These KF-25 strain-specific genes mainly exhibited high levels of similarity to genes from *Aspergillus* species and *Neosartorya fischeri* (Additional file 1: Figure S4), which are evolutionarily closely related to *P. chrysogenum* [12]. The biological functions of the 355 KF-25-specific genes are mainly unknown (232, representing 65%), except for some ORFs that were predicted to be involved in transport, metabolism, and signal transduction (Figure 3c; Additional file 1: Table S6 and Figure S3B). Among the 355 ORFs, none were found to be involved in cell mobility, extracellular structures, chromatin

structure, and metabolism (Figures 2 and 3c), but ORFs with functions in intracellular trafficking, secretion, vesicular transport, signal transduction, and transcription were frequently found. To confirm that the 2.3 Mb of DNA fragments were truly missing from the KF-25 genome, three randomly chosen Wisconsin 54-1255 strain-specific genes (Pc03g00290, Pc12g02270 and Pc21g20980) from these fragments of were investigated using PCR amplification. The results showed that these genes were detected in the Wisconsin 54-1255 genome but not in the KF-25 genome (Figure 4). Another one Wisconsin 54-1255-specific gene, Pc00c02 [GenBank:AM920417.1], which is annotated as a 16S ribosomal RNA was not

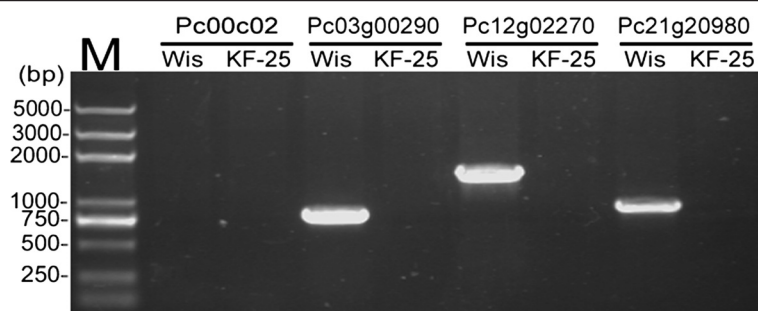


Figure 4 PCR detection of the *P. chrysogenum* Wisconsin 54-1255-specific genes from the genome of KF-25 and Wisconsin 54-1255.

detected in either genome. The 16S ribosomal RNA gene is widely used to classify bacteria and is reported to only exist in bacterial genomes [40]. BLAST analysis of Pc00c02 indicated that it was highly similar to the 16S rDNA region of bacteria *Rugosimonospora* sp. 260305 (100% identify) and *Micromonospora* sp. HBUM80369 (99% identify). The 16S rDNA found in the Wisconsin 54-1255 genome sequence might be caused by bacterial contamination during sequencing.

Comparative analysis of *P. chrysogenum* KF-25 and other *P. chrysogenum* strains

According to previous proteomic studies, the improvement process of penicillin production enhanced the expression of some genes, while decreasing [15,41,42]. *P. chrysogenum* Wisconsin 54-1255 is a moderately improved penicillin producer derived from the wild-type *P. chrysogenum* NRRL 1951, which exhibited more secondary metabolism pathways (such as pigments), pathogenicity proteins and virulence proteins compared with Wisconsin 54-1255 and another high penicillin producer *P. chrysogenum* AS-P-78 [15,42]. *P. chrysogenum* KF-25 is a wild-type strain that had a stronger yellow pigment production than Wisconsin 54-1255 (Figure 1). The ability to produce more pigments is representative of a greater number of secondary metabolic pathways, and was a common feature of both KF-25 and NRRL 1951. Several KF-25-specific genes were found to be associated with pathogenicity and virulence. One such gene, KF25_6369, which encodes glucose oxidase, is thought to be involved in virulence because gluconic acid and glucose oxidase are related to pathogenicity of *Penicillium expansum* in apples [43]. Glucose oxidase also showed reduced expression in Wisconsin 54-1255, compared with NRRL 1951 [42]. The penicillin synthesis genes were clustered in one group in the genomes of NRRL 1951 and Wisconsin 54-1255, while several such clusters were found in the AS-P-78 genome [44]. Similar to wild-type NRRL 1951, KF-25 contained only one penicillin synthesis gene cluster. Wild-type *P. chrysogenum* KF-25 and NRRL 1951 have more

secondary metabolism pathways and more pathogenicity and virulence associated genes, which are fitness mechanisms for the wild-type strains to survive in natural environment.

Phylogenetic analysis of *P. chrysogenum* KF-25 and the other sequenced filamentous fungi

A concatenated set of the amino acid sequences of 90 conserved proteins was used to construct a phylogenetic tree [12]. The phylogenetic analysis (Figure 5) showed a close relationship between KF-25 and *Aspergillus* species, and a more distant evolutionary relationship between KF-25 and *Penicillium marneffeii* and *Talaromyces stipitatus*. *P. chrysogenum* KF-25 was in the same evolutionary branch as Wisconsin 54-1255 and showed a close relationship with *Penicillium digitatum*. This result was consistent with previous reports [12,45]. A phylogenetic tree constructed based on the amino acid sequences of the β -tubulin also supported the evolutionary relationship of strains from the *Penicillium* genus (Additional file 1: Figure S5).

Secondary metabolism analysis of *P. chrysogenum* KF-25 Putative secondary metabolism pathways

P. chrysogenum has been known as a penicillin producer for many years [1]. Recently, studies have mainly focused on the pathways of penicillin synthesis, and the key genes involved in penicillin production have been determined [13,27,46]. In addition to penicillin, *P. chrysogenum* can produce many other secondary metabolites, such as mycotoxin and drugs [27,28,47,48]. In a previous report, SMURF analysis predicted that the genome of Wisconsin 54-1255 contains 33 secondary metabolism gene clusters [49]. In this study, secondary metabolism gene clusters was predicted using anti-SMASH [29], and 33 and 41 gene clusters were identified in the genomes of KF-25 and Wisconsin 54-1255 (Additional file 1: Table S7 and Figure S6). The predicted products of 23 secondary metabolism gene clusters in KF-25 were: eight nonribosomal peptides, 10 polyketides, two hybrid non-ribosomal peptide synthase

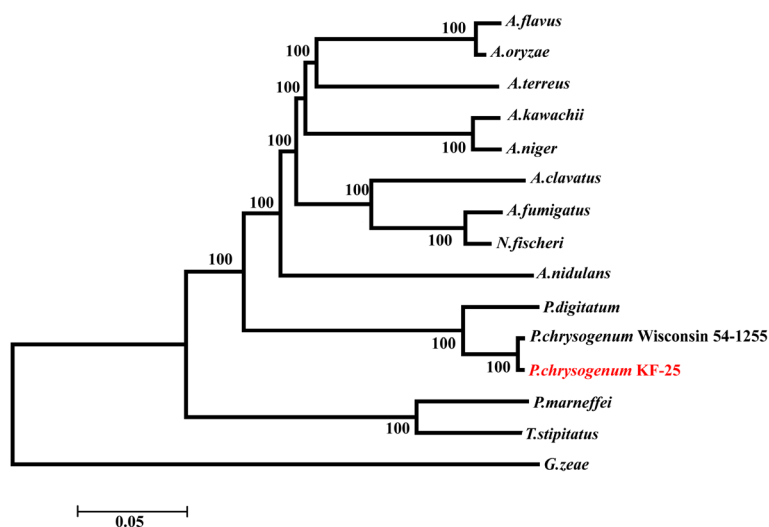


Figure 5 Phylogenetic analysis of *P. chrysogenum* and other sequenced filamentous fungi. The phylogenetic tree was constructed by concatenating the amino acid sequences of 90 conserved proteins from all strains using the neighbor-joining method and bootstrap analysis (1,000 replicates) of the ClustalW alignment. *Gibberella zeae* was used as an outgroup strain. The support rates are indicated at the node of each branch and the scale bar represents 0.05 substitutions per amino acid position.

(NRPS)-polyketide synthases (PKS), one hybrid NRPS-terpene, one terpene and one siderophore, while the remaining 10 gene clusters produced other secondary metabolites (Additional file 1: Table S7). Among the 33 gene clusters, five were predicted to produce stigmatellin, chalcomycin, epothilone, fumitremorgin and penicillin. The production of penicillin by KF-25 and Wisconsin 54-1255 were verified by HPLC (Additional file 1: Figure S7). The data showed that Wisconsin 54-1255 exhibited greater ability of producing penicillin than KF-25.

Non-ribosomal peptide synthetase

NRPSs play important roles in the synthesis of non-ribosomal peptides, which include antibiotics and other important pharmaceuticals [50]. In the *P. chrysogenum* KF-25 genome, 20 NRPS genes were found and the domain compositions of these predicted NRPSs are shown in Additional file 1: Figure S8. Among the 20 predicted NRPSs, 14 were involved in putative secondary metabolism pathways, while the other six were not. Eleven of the 20 predicted NRPSs, encoded by gene KF25_6155, KF25_1342, KF25_6525, KF25_1526, KF25_9456, KF25_5703, KF25_8966, KF25_6509, KF25_8398, KF25_9347, and KF25_4993, had similar amino acid sequences to HC-toxin synthase. In addition, 15 of the predicted MFS transporters encoded by the KF-25 genome were identified as HC-toxin efflux carriers. The HC-toxins determine the specificity and virulence of pathogenic fungi toward host plants [51]. The existence of the HC-toxin synthases and HC-toxin efflux carriers suggested that *P. chrysogenum* KF-25 might produce HC-toxin.

Polyketide synthase

Polyketides, including pigments, antibiotics, and mycotoxins, are a diverse group of secondary metabolites produced by microorganisms and plants. PKSs are complex enzymatic systems for producing polyketides [52-55]. Type I and type II PKSs are modular in structure and contain multiple catalytic activity enzymes individually [56], while the type III PKSs have simple structures [57,58]. There were 10 predicted polyketide synthesis pathways and two predicted hybrid NRPS-PKS synthesis pathways in the KF-25 genome sequence. Twenty-four polyketide synthase genes were extracted from the KF-25 genome and 23 of them were predicted to encode type I PKSs. The remaining gene (KF25_7297) encoded a type III PKS (Figure 6a). Thirteen of 24 predicted PKSs were identified as members of putative secondary metabolism pathways. One such pathway, containing a type I PKS was predicted to produce epothilone (Additional file 1: Table S7). According to previous reports, epothilone is produced by myxobacteria and exhibits anticancer activity by targeting the microtubule of the cancer cell [59]. Because the KF-25 genome contains an epothilone synthesis gene cluster, it is possible that KF-25 might be useful in producing this potential anticancer agent. We will further investigate whether KF-25 produces epothilone and whether the strain has anticancer activity. Type I PKSs have similarity to the type-I fatty acid synthases (FAS), which are essential in lipid metabolism [55,56]. The existence of diverse PKS genes in the *P. chrysogenum* KF-25 genome suggests that KF-25 might produce diverse lipids and polyketides, and that

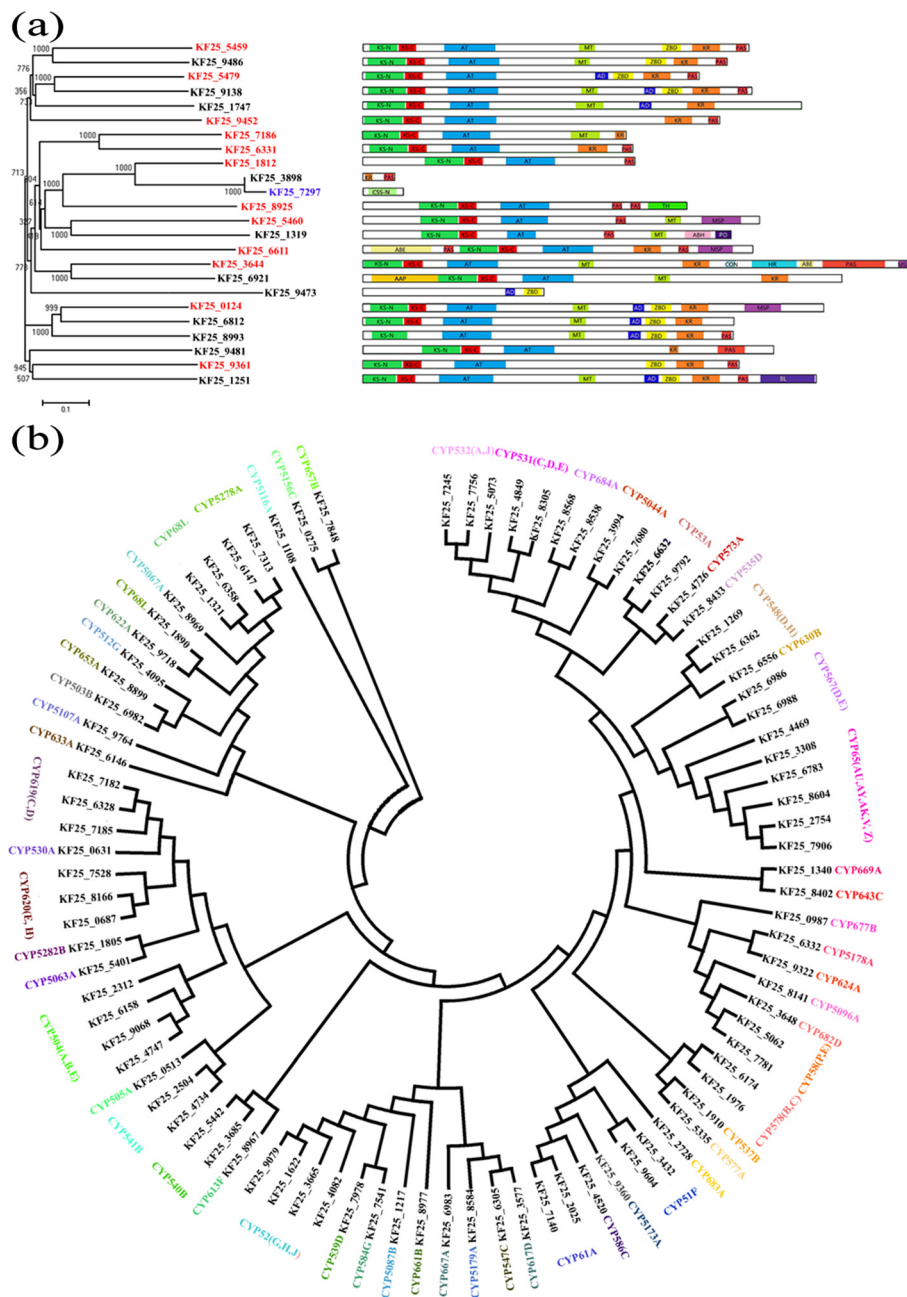


Figure 6 Neighbor-joining phylogenetic tree of polyketide synthases (PKSs) and cytochrome P450 from the KF-25 genome. The amino acid sequences of the proteins were used to construct the phylogenetic tree using ClustalX2.0 with the neighbor-joining method. The branch length scale bar below the phylogenetic tree indicates the number of substitutions per amino acid site. **(a)** The functional domain architecture of proteins was predicted using Pfam and AMSPKS [89,90] and is shown on the right. Protein domain names were as follows: KS_N, β -ketoacyl synthase, N-terminal domain; KS_C, β -ketoacyl synthase, N-terminal domain; AT, acyl transferase; KR, β -keto reductase; PAS, phosphopantetheine attachment site; MT, methyltransferase; MSP, male sterility protein; TH, thioesterase; ZBD, zinc-binding dehydrogenase; ER, ER domain; AD, alcohol dehydrogenase GroES-like domain; CSS_N, chalcone and stilbene synthases, N-terminal domain; ABH, α/β hydrolase; PO, prolyl oligopeptidase; CON, condensation domain; HR, HxxPF-repeated domain; ABE, AMP-binding enzyme; BL, β -lactamase; AAP, amino acid permease. The ORFs indicated in red (type I PKS) and blue (type II PKS) are the members of the putative secondary metabolism pathways. **(b)** The cytochrome P450 (CYPs) identified in the KF-25 genome and CYPs of different families are indicated in different colors. The families of the corresponding CYPs are indicated beside the name of the proteins.

these metabolic products might influence the life cycle of *P. chrysogenum*.

Cytochrome P450

Cytochrome P450s (CYPs) are hemoproteins that are ubiquitously distributed throughout all domains of life and play important and diverse roles in metabolic processes and adaptation to different environmental niches by fungi [60]. CYPs participating in numerous primary, secondary, and xenobiotic metabolic reactions have been reported [61,62], and several CYPs predicted from sequenced microorganism genomes were found to be members of secondary metabolism pathways [63,64]. CYPs can be classified into different families based on the amino acid sequences [65,66]. Ninety CYPs were predicted in the KF-25 genome (about 0.9% of total ORFs) and many of them were members of putative secondary metabolism pathways, including the pathways of PKSs, NRPSs, and NRPS-terpenes. These CYPs belonged to 60 different families. There were usually one or two CYPs per family but some families contained three to six CYPs (Figure 6b). The classifications of the CYPs from the Wisconsin 54-1255 genome were almost the same as those from KF-25 genome (Additional file 1: Figure S9). As a multicomponent electron transport chain system, CYPs are critical in degradation, detoxification, and syntheses of life-critical compounds in organisms [67]. Besides their functions in secondary metabolism, CYPs also play critical roles in the adaptation of organisms to specific ecological niches and the biosynthesis of physiologically important compounds [68,69]. The existence of so many CYPs might be essential for the life cycle *P. chrysogenum* and the synthesis of the metabolic products, such as penicillin [70].

P. chrysogenum virus terminal fragment-similar sequences

To date, the genome of only one virus originating from *P. chrysogenum* has been sequenced, which showed it was a dsRNA virus of the *Chrysovirus* genus [71,72]. DNA alignment analysis (Figure 7) showed that numerous sequences in the KF-25 genome were similar to the 5'- and 3'-UTR of four *P. chrysogenum* virus DNA sequence segments [72]. These sequences were also found in the genome of Wisconsin 54-1255 (data not shown). The sequences matching the 5'-UTR of the virus were mainly composed of (CAA)_n repeats, which are similar to the translational enhancer elements in the 5'-UTR of tobacco viruses [73]. Some sequence fragments of KF-25 genome matched the 5'-UTRs and 3'-UTR of virus segment 2, but did not contain regions encoding virus structural proteins. According to previous reports, eukaryotic genomes contain many sequence of viral origin that have played diverse roles, such as horizontal gene transfer mediated by dsRNA viruses, providing resistance to the virus,

and promoting the evolution of host organisms [74-77]. We speculate that *P. chrysogenum* genome might have obtained the UTRs by integrating the viral genome. During the evolutionary process, genes encoding virus structural proteins were eliminated but the UTR regions remained. The functions of these sequences in *P. chrysogenum* genomes are still unknown, but they might provide insertion sites for the virus, or a potential mechanism of viral resistance for *P. chrysogenum*.

Conclusions

In this study, we reported the genome sequence of wild-type *P. chrysogenum* KF-25. This is the second report of a *P. chrysogenum* genome, but the first of wild-type strain. Comparative genome analysis showed that KF-25 genome lacked regions of the genome, totaling 2.3 Mb, that were found in Wisconsin 54-1255 genome, which were previously considered to be *P. chrysogenum* species-specific regions [12]. However, our results showed that the missing regions were only specific to Wisconsin 54-1255. These regions contained numerous repeat elements and transposable elements, indicating that these segments might have been obtained by Wisconsin 54-1255 through transposition and horizontal gene transfer during evolution. Comparative analysis of KF-25 with another wild-type strain, NRRL 1951, revealed that they had numerous features in common, such as pigments production, and a greater number of pathogenicity- and virulence-associated genes. Based on the phylogenetic tree of 90 conserved orthologous proteins, strains KF-25 and Wisconsin 54-1255 maintained a close evolutionary distance. Analysis of the TCRs indicated that many proteins were osmolarity TCRs, which may be an adaptive strategy of *P. chrysogenum* to high osmotic pressure. Several gene clusters involved in putative secondary metabolism pathways, and many genes encoding essential enzymes for the biosynthesis of diverse biologically-active agents were found, which could provide foundation for using *P. chrysogenum* to produce antibiotics including penicillin and other β -lactam antibiotics. The identification of *P. chrysogenum* virus UTR sequences in the two sequenced *P. chrysogenum* genomes is helpful for studying the relationship between the virus and its fungal host in evolution. The results of this study can help us to further understand the genetic diversity of *P. chrysogenum* and shed light on its evolution, biology, environmental adaptation and application.

Methods

Strains and culture conditions

P. chrysogenum strain KF-25 and *U. virens* strain UV-1 were isolated and identified by our lab. Strain Wisconsin 54-1255 [12] was provided by MA van den Berg at DSM Anti-Infectives. Fungal strains were grown in potato-

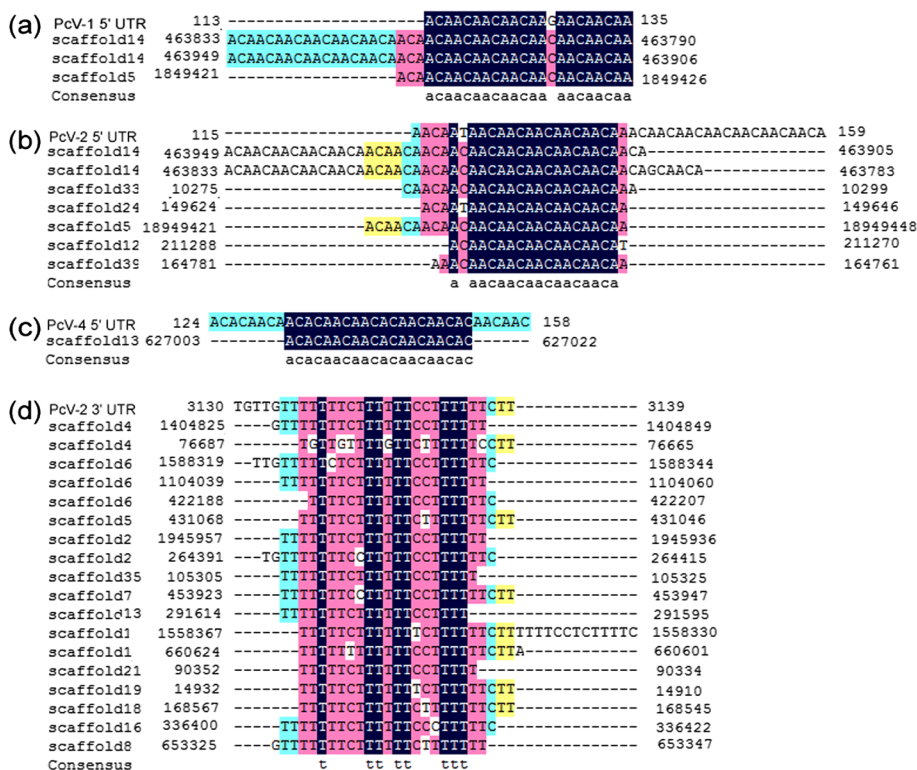


Figure 7 Sequence alignments of segments from the *P.chrysogenum* KF-25 genome with the 3'-and 5'-UTRs of *P. chrysogenum* virus. The sequences of three segments from *P. chrysogenum* dsRNA virus are aligned with the genome of KF-25, showing the sequence similarity of the segments from the KF-25 genome to the 3'-and 5'-UTR of the three dsRNA virus segments. Sequences exhibiting similarity to the PcV-1 5'-UTR (a), the PcV-2 5'-UTR (b), the PcV-4 5'-UTR (c) and the PcV-2 3'-UTR (d) were indicated.

sucrose (PS) medium [20% (w/v) potato lixivium, 2% (w/v) sucrose], and 1.5% (w/v) agar was used in solid potato-sucrose medium (PSA). To assay the antifungal activity, *P. chrysogenum* strains KF-25 and Wisconsin 54-1255 were grown in 500-ml flasks containing 100 ml of PS medium at 28°C for 96 h with shaking (180 rpm). The culture supernatants were filtered through four layers of cheesecloth and centrifugated at 16000 × g for 20 min at 4°C. The culture supernatants were sterilized by filtering through a 0.22 μm membrane (Millipore) and were used to assay the antifungal activity against *U. virens* using the disk diffusion test [78]. The conidia of pathogen *U. virens* were spread on a PSA plate at a density of 10⁸ spores/ml and 100 μl spore suspension was used for each plate, then 20 μl of the sterilized culture supernatant above was added to a piece of sterile filter paper with a 6 mm diameter, placed in the center of the plate. The plate was incubated for 5 days at 28°C. Assays were performed in triplicate.

HPLC-DAD analysis

Conidiospores of *P. chrysogenum* KF-25 and Wisconsin 54-1255 were inoculated at 10⁵ to 10⁶ conidia/ml in a production medium containing (g/l): glucose · H₂O, 5;

lactose · H₂O, 80; (NH₂)₂CO, 4.5; (NH₄)₂SO₄, 1.1; Na₂SO₄, 2.9; KH₂PO₄, 5.2; K₂HPO₄ · 3H₂O, 4.8; trace elements solution (citric acid · H₂O, 150; FeSO₄ · 7H₂O, 15; MgSO₄ · 7H₂O, 150; H₃BO₃, 0.0075; CuSO₄ · 5H₂O, 0.24; CoSO₄ · 7H₂O, 0.375; ZnSO₄ · 7H₂O, 1.5; MnSO₄ · H₂O, 2.28; CaCl₂ · 2H₂O, 0.99), 10 (ml/l); 10% phenylacetic acid, pH 7, 75 (ml/l) and the pH was adjusted to pH 6.5 before inoculation [79]. The culture was incubated at 25°C in an orbital shaker at 280 rpm for 4 days. The mycelium was removed by centrifugation and filtration, and the fermentation broth was assayed for penicillin by HPLC-DAD (High Performance Liquid Chromatography-Diode Array Detector). The assay was performed on a Dionex UltiMate 3000 RS HPLC system with autosampler and a DAD detector (Thermo Fisher Scientific, Waltham, MA) and an Agilent ZORBAX 300SB-C18 column (250 × 4.6 mm, 5 μm particle size, Agilent Technologies, Palo Alto, CA). The mobile phase was consisted of solvents A [0.5 mol/L KH₂PO₄ (pH3.5): methanol: water, 1:3:6] and B [0.5 mol/L KH₂PO₄ (pH3.5): methanol: water, 1:5:4]. The gradient program started with 30% of B, followed by increasing to 100% B from 0 to 20 min, held at 100% B from 20 to 35 min, decreasing to 30% of B from 35 to 50 min. The flow rate was 1.0 ml/min with a column temperature of

25°C. The injection volume was 20 µl, and the detection wavelength was 210 nm. Penicillin G (0.5 mg/ml) was used as a positive control.

The cultures of *P. chrysogenum* KF-25 and Wisconsin 54-1255 in potato-sucrose (PS) medium for 4 days were analysed by HPLC on a Dionex UltiMate 3000 RS HPLC system with autosampler and a DAD detector and a Sepax Polar-Silica column (250 × 10.0 mm, 5 µm particle size, Sepax Technologies, Newark, DE). The mobile phase consisted of solvents A (10 mM ammonium acetate) and B (methanol). The program held at 80% B from 0 to 20 min. The flow rate was 2.0 ml/min and the column temperature was 25°C. The injection volume was 5 µl, and the detection wavelength was 210 nm.

Genome sequencing, assembly, and annotation

Whole-genome sequencing of KF-25 was performed by the National Center for Gene Research, Shanghai, China. KF-25 genomic DNA was extracted as described previously [80], then was randomly sheared and purified to construct three libraries with insert sizes of 170 bp, 500 bp and 2–3 kb. DNA was amplified from the libraries and sequenced by HiSeq2000 (Illumina, California, USA). The reads were assembled into contigs by Velvet (Version 1.2.03) [81] and then scaffolds were constructed based on the contigs using SSPACE [82].

AUGUSTUS (http://bioinf.uni-greifswald.de/augustus/) [83] was used to predict the genes in the KF-25 genome, and the putative proteins were aligned against the NCBI nr, UniProt (http://www.uniprot.org/) and KEGG (http://www.genome.jp/kegg/) database using BLASTP tool (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The predicted genes were then aligned against the CDD database (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) using rpsBLAST. To identify the KOG classification of each gene, we searched for each amino acid sequence in the KOG database in NCBI using KOGnitor (http://www.ncbi.nlm.nih.gov/COG/grace/kognitor.html). Metabolic pathways of the KF-25 genome were constructed based on the annotation results against the KEGG database. Repeat sequences were analyzed using CENSOR [84] (http://www.girinst.org/censor/index.php). The genes encoding tRNA were predicted using tRNAScan [85], and RNAmmer [40] was used to find rDNA sequences. *P. chrysogenum* virus terminal UTR sequences in the KF-25 genome were identified using local BLAST and the sequences were aligned using ClustalX 2.0 [86].

Comparative genome analysis

Mauve software [87] was used to compare the genome of KF-25 with Wisconsin 54-1255 [GenBank:NS_000201.1]. Dot plot analysis of the two genomes was performed with Gepard [88]. The orthologous genes between KF-25 and Wisconsin54-1255 were by compared the proteomes of

the two genomes and proteins that exhibited similarity higher than 25% were thought orthologous. Proteins encoded by all of the strain-specific genes were classified by searching the eukaryotic orthologous groups (KOG) database in NCBI using KOGnitor.

Detection of strain-specific genes from *P. chrysogenum*

The genomic DNA of KF-25 and Wisconsin 54-1255 was extracted as described previously [80]. Four pairs of primers based on specific gene sequences of Wisconsin 54-1255 (Additional file 1: Table S8) were used to amplify specific genes by PCR (primers used were listed in Additional file 1: Table S8). The products were detected on an agarose gel.

Secondary metabolism-related gene analysis

The secondary metabolism pathways in the KF-25 and Wisconsin 54-1255 genomes were predicted using antiSMASH (http://antismash.secondarymetabolites.org/) [29]. Modular polyketide synthases in the genome were predicted and the domain compositions were analyzed using AMSPKS [89] and Pfam [90]. Genome-encoded cytochrome P450s were classified by searching against the fungal cytochrome P450 database [65].

Phylogenetic analysis

Phylogenetic trees were constructed in MEGA 5.05 [91], using the neighbor-joining method and bootstrap analysis (1,000 replicates), of MUSCLE [92] or ClustalW [86] alignments. Phylogenetic trees of filamentous fungi were constructed as described previously [12] using the aligned amino acid sequences of 90 orthologous genes from *P. chrysogenum* KF-25, *P. chrysogenum* Wisconsin 54-1255 [GenBank:NS_000201.1], *P. marneffeii* [GenBank:ABAR000000000], *P. digitatum* [45], *T. stipitatus* [GenBank:ABAS000000000], *A. niger* [93], *A. nidulans* [GenBank:AA CD000000000], *A. oryzae* [94], *Aspergillus fumigatus* [95], *Aspergillus clavatus* [GenBank:AAKD000000000], *Aspergillus terreus* [GenBank:AAJN000000000], *Aspergillus flavus* [GenBank:AAIH000000000], *Aspergillus kawachii* [96], *Neosartorya fischeri* [GenBank:AAKE000000000], and *Gibberella zeae* [GenBank:AACM000000000] (Additional file 1: Table S9).

Data access

The complete genome sequence of *P. chrysogenum* KF-25 has been submitted to SRA (http://www.ncbi.nlm.nih.gov/sra/) under the accession number SRP022930.

Additional file

Additional file 1: Table S1. Anticodon usage of *Penicillium chrysogenum* KF-25 genome. **Figure S1.** Number of occurrences of simple sequence repeats in *P. chrysogenum* KF-25 genome. **Table S2.**

Putative transcription factors in the genome of *P. chrysogenum* KF-25. **Table S3.** Putative translation factors in the *P. chrysogenum* KF-25 genome. **Table S4.** List of the ORFs with the predicted function as the compositions of the secretion system. **Figure S2.** Dot plot analysis of *P. chrysogenum* KF-25 (horizontal) and *P. chrysogenum* Wisconsin 54–1255 (vertical) genomes. **Table S5.** KOG annotation of the *P. chrysogenum* Wisconsin 54–1255 specific ORFs. **Figure S3.** Functional classification of the *P. chrysogenum* Wisconsin 54–1255 and KF-25 specific ORFs based on the KOG database. **Table S6.** KOG annotation of the *P. chrysogenum* KF-25 specific ORFs. **Figure S4.** Classifications of the origin of the most similar genes in GenBank of the 355 KF-25 specific genes. **Figure S5.** Neighbor-Joining phylogenetic tree of *P. chrysogenum* KF-25 and other species of the genus of *penicillium* based on the *benA* gene. **Table S7.** Detail information of the predicted secondary metabolism gene clusters. **Figure S6.** Putative structures of the predicted secondary metabolism gene clusters products. **Figure S7.** Detection of penicillin G by HPLC-DAD. **Figure S8.** The domain compositions and the phylogenetic tree of the non-ribosomal synthetases from KF-25 genome. **Figure S9.** Neighbor-Joining (NJ) phylogenetic tree of the cytochrome P450 (CYPs) from the genomes of *P. chrysogenum* KF-25 and *P. chrysogenum* Wisconsin 54–1255. **Table S8.** Primers used to amplify the *P. chrysogenum* Wisconsin 54–1255 specific genes from both the genomes of *P. chrysogenum* Wisconsin 54–1255 and *P. chrysogenum* KF-25. **Table S9.** Orthologous genes used in phylogenetic analysis of various filamentous fungi.

Competing interests

The authors have no competing interests to declare.

Authors' contributions

MYG, QP and YHY designed the study. QP, YHY, XPC and BL performed the experiments and analyzed the data. QP and YHY drafted the manuscript. MYG revised the manuscript. PML, YW and DDW provided reagents. All authors reviewed and approved the final version of the manuscript.

Acknowledgement

We thank van den Berg MA at DSM Anti-Infectives for providing *Penicillium chrysogenum* Wisconsin 54-1255. This study was supported by the National Natural Science Foundation of China (No.31170123, 31201560), the National Project (2009ZX08009-056B), and the projects of the Chinese Academy of Sciences (KSCX2-EW-G-16).

Author details

¹Key Laboratory of Agricultural and Environmental Microbiology, Wuhan Institute of Virology, Chinese Academy of Sciences, Wuhan 430071, China. ²Present address: Hubei University of Economics, Wuhan, Hubei, China.

Received: 31 July 2013 Accepted: 6 February 2014

Published: 21 February 2014

References

1. Fleming A: On the Antibacterial Action of Cultures of a *Penicillium*, with Special Reference to Their Use in the Isolation of *B. Influenzae*. *Br J Exp Pathol* 1929, **10**(3):226–236.
2. Satyavathi K, Kiran TN, Raju PB, Sharmila M: Synthesis and Screening of 3-Formyl-2-thio-1,2,3,4-tetrahydro pyrimidine Analogues as Antibacterial Agents. *Asian J Chem* 2010, **22**(7):5182–5186.
3. Deng XH, Chai LY, Yang ZH, Tang CJ, Tong HX, Yuan PF: Bioleaching of heavy metals from a contaminated soil using indigenous *Penicillium chrysogenum* strain F1. *J Hazard Mater* 2012, **233**:25–32.
4. Ferreira-Guedes S, Mendes B, Leitao AL: Degradation of 2,4-dichlorophenoxyacetic acid by a halotolerant strain of *Penicillium chrysogenum*: antibiotic production. *Environ Technol* 2012, **33**(6):677–686.
5. Kumar S, Mathur A, Singh V, Nandy S, Khare SK, Negi S: Bioremediation of waste cooking oil using a novel lipase produced by *Penicillium chrysogenum* SNPs grown in solid medium containing waste grease. *Bioresour Technol* 2012, **120**:300–304.
6. Murali M, Sudisha J, Amruthesh KN, Ito SI, Shetty HS: Rhizosphere fungus *Penicillium chrysogenum* promotes growth and induces defence-related genes and downy mildew disease resistance in pearl millet. *Plant Biol* 2013, **15**(1):111–118.
7. Leiter E, Szappanos H, Oberparleiter C, Kaiserer L, Csernoch L, Pusztahelyi T, Emri T, Pocsí I, Salvenmoser W, Marx F: Antifungal protein PAF severely affects the integrity of the plasma membrane of *Aspergillus nidulans* and induces an apoptosis-like phenotype. *Antimicrob Agents Chemother* 2005, **49**(6):2445–2453.
8. Rodriguez-Martin A, Acosta R, Liddell S, Nunez F, Benito MJ, Asensio MA: Characterization of the novel antifungal protein PgAFP and the encoding gene of *Penicillium chrysogenum*. *Peptides* 2010, **31**(4):541–547.
9. Rodriguez-Martin A, Acosta R, Liddell S, Nunez F, Benito MJ, Asensio MA: Characterization of the novel antifungal chitosanase PgChP and the encoding gene from *Penicillium chrysogenum*. *Appl Microbiol Biotechnol* 2010, **88**(2):519–528.
10. Marx F, Binder U, Leiter E, Pocsí I: The *Penicillium chrysogenum* antifungal protein PAF, a promising tool for the development of new antifungal therapies and fungal cell biology studies. *Cell Mol Life Sci* 2008, **65**(3):445–454.
11. Hegedus N, Leiter E, Kovacs B, Tomori V, Kwon NJ, Emri T, Marx F, Batta G, Csernoch L, Haas H, Yu JH, Pocsí I: The small molecular mass antifungal protein of *Penicillium chrysogenum* - a mechanism of action oriented review. *J Basic Microbiol* 2011, **51**(6):561–571.
12. van den Berg MA, Albang R, Albermann K, Badger JH, Daran JM, Driessen AJM, Garcia-Estrada C, Fedorova ND, Harris DM, Heijne WHM, Joardar V, Kiel JAKW, Kovalchuk A, Martin JF, Nieman WC, Nijland JG, Pronk JT, Roubos JA, van der Klei IJ, van Peij NNME, Veenhuis M, von Dohren H, Wagner C, Wortman J, Bovenberg RAL: Genome sequencing and analysis of the filamentous fungus *Penicillium chrysogenum*. *Nat Biotechnol* 2008, **26**(10):1161–1168.
13. van den Berg MA: Impact of the *Penicillium chrysogenum* genome on industrial production of metabolites. *Appl Microbiol Biotechnol* 2011, **92**(1):45–53.
14. Fassatio O: Contribution to Morphology of Productive Strains of *Penicillium-Chrysogenum* Thom from Wisconsin Family. *Folia Microbiol* 1970, **15**(5):358.
15. Jami MS, Barreiro C, Garcia-Estrada C, Martin JF: Proteome Analysis of the Penicillin Producer *Penicillium chrysogenum* characterization of protein changes during the industrial strain improvement. *Mol Cell Proteomics* 2010, **9**(6):1182–1198.
16. Rodriguez-Saiz M, Barredo JL, Moreno MA, Fernandez-Canon JM, Penalva MA, Diez B: Reduced function of a phenylacetate-oxidizing cytochrome P450 caused strong genetic improvement in early phylogeny of penicillin-producing strains. *J Bacteriol* 2001, **183**(19):5465–5471.
17. Scott J, Untereiner WA, Wong B, Straus NA, Malloch D: Genotypic variation in *Penicillium chrysogenum* from indoor environments. *Mycologia* 2004, **96**(5):1095–1105.
18. Henk DA, Eagle CE, Brown K, Van den Berg MA, Dyer PS, Peterson SW, Fisher MC: Speciation despite globally overlapping distributions in *Penicillium chrysogenum*: the population genetics of Alexander Fleming's lucky fungus. *Mol Ecol* 2011, **20**(20):4288–4301.
19. Kim KW, Park EW: Ultrastructure of spined conidia and hyphae of the rice false smut fungus *Ustilaginoidea virens*. *Micron* 2007, **38**(6):626–631.
20. Willi Y, Maattanen K: Evolutionary dynamics of mating system shifts in *Arabidopsis lyrata*. *J Evol Biol* 2010, **23**(10):2123–2131.
21. Blouin MS, Parsons M, Lacaillie V, Lotz S: Use of microsatellite loci to classify individuals by relatedness. *Mol Ecol* 1996, **5**(3):393–401.
22. Huang QY, Xu FH, Shen H, Deng HY, Liu YJ, Liu YZ, Li JL, Recker RR, Deng HW: Mutation patterns at dinucleotide microsatellite loci in humans. *Am J Hum Genet* 2002, **70**(3):625–634.
23. Pallen MJ, Chaudhuri RR, Henderson IR: Genomic analysis of secretion systems. *Curr Opin Microbiol* 2003, **6**(5):519–527.
24. Cao TB, Saier MH: The general protein secretory pathway: phylogenetic analyses leading to evolutionary conclusions. *Biomembranes* 2003, **16**(9):115–125.
25. Gowda K, Black SD, Moller I, Sakakibara Y, Liu MC, Zwieb C: Protein SRP54 of human signal recognition particle: cloning, expression, and comparative analysis of functional sites. *Gene* 1998, **207**(2):197–207.
26. Fagan RP, Fairweather NF: *Clostridium difficile* Has Two Parallel and Essential Sec Secretion Systems. *J Biol Chem* 2011, **286**(31):27483–27493.
27. Gidijala L, Kiel JAKW, Bovenberg RAL, Van Der Klei IJ, Van Den Berg MA: Biosynthesis of active pharmaceuticals: beta-lactam biosynthesis in filamentous fungi. *Biotechnol Bioeng* 2010, **27**:1–31.
28. Kozlovskii AG, Zhelifonova VP, Adanin VM, Antipova TV, Shnyreva AV, Viktorov AN: The biosynthesis of low-molecular-weight nitrogen-containing secondary metabolites-alkaloids-by the resident strains of *Penicillium chrysogenum* and *Penicillium expansum* isolated on board the Mir space station. *Microbiology* 2002, **71**(6):666–669.

29. Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, Weber T, Takano E, Breitling R: **antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences.** *Nucleic Acids Res* 2011, **39**:W339–W346.
30. Yang J, Xu XX, Liu G: **Amplification of an MFS Transporter Encoding Gene penT Significantly Stimulates Penicillin Production and Enhances the Sensitivity of *Penicillium chrysogenum* to Phenylacetic Acid.** *J Genet Genomics* 2012, **39**(11):593–602.
31. Weber SS, Kovalchuk A, Bovenberg RAL, Driessen AJM: **The ABC transporter ABC40 encodes a phenylacetic acid export system in *Penicillium chrysogenum*.** *Fungal Genet Biol* 2012, **49**(11):915–921.
32. Catlett NL, Yoder OC, Turgeon BG: **Whole-genome analysis of two-component signal transduction genes in fungal pathogens.** *Eukaryot Cell* 2003, **2**(6):1151–1161.
33. Furukawa K, Hoshi Y, Maeda T, Nakajima T, Abe K: ***Aspergillus nidulans* HOG pathway is activated only by two-component signalling pathway in response to osmotic stress.** *Mol Microbiol* 2005, **56**(5):1246–1261.
34. Wolanin PM, Thomason PA, Stock JB: **Histidine protein kinases: key signal transducers outside the animal kingdom.** *Genome Biol* 2002, **3**(10):REVIEWS3013.
35. Chang C, Stewart RC: **The two-component system - Regulation of diverse signaling pathways in prokaryotes and eukaryotes.** *Plant Physiol* 1998, **117**(3):723–731.
36. Appleyard MVCL, McPheat WL, Stark MJR: **A novel 'two-component' protein containing histidine kinase and response regulator domains required for sporulation in *Aspergillus nidulans*.** *Curr Genet* 2000, **37**(6):364–372.
37. Duran R, Cary JW, Calvo AM: **Role of the Osmotic Stress Regulatory Pathway in Morphogenesis and Secondary Metabolism in Filamentous Fungi.** *Toxins* 2010, **2**(4):367–381.
38. Mert HH, Ekmekci S: **The Effect of Salinity and Osmotic-Pressure of the Medium on the Growth, Sporulation and Changes in the Total Organic-Acid Content of *Aspergillus-Flavus* and *Penicillium-Chrysogenum*.** *Mycopathologia* 1987, **100**(2):85–89.
39. Luard EJ: **Respiration of *Penicillium-Chrysogenum* in Relation to the Osmotic Potential of the Growth-Medium.** *Exp Mycol* 1985, **9**(2):99–107.
40. Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW: **RNAmer: consistent and rapid annotation of ribosomal RNA genes.** *Nucleic Acids Res* 2007, **35**(9):3100–3108.
41. Jami MS, Garcia-Estrada C, Barreiro C, Cuadrado AA, Salehi-Najafabadi Z, Martin JF: **The *Penicillium chrysogenum* Extracellular Proteome. Conversion from a Food-rotting Strain to a Versatile Cell Factory for White Biotechnology.** *Mol Cell Proteomics* 2010, **9**(12):2729–2744.
42. Barreiro C, Martin JF, Garcia-Estrada C: **Proteomics Shows New Faces for the Old Penicillin Producer *Penicillium chrysogenum*.** *J Biomed Biotechnol* 2012, **2012**:105109.
43. Hadas Y, Goldberg I, Pines O, Prusky D: **Involvement of gluconic acid and glucose oxidase in the pathogenicity of *Penicillium expansum* in apples.** *Phytopathology* 2007, **97**(3):384–390.
44. Fierro F, Barredo JL, Diez B, Gutierrez S, Fernandez FJ, Martin JF: **The Penicillin Gene-Cluster Is Amplified in Tandem Repeats Linked by Conserved Hexanucleotide Sequences.** *Proc Natl Acad Sci U S A* 1995, **92**(13):6200–6204.
45. Marcet-Houben M, Ballester AR, de la Fuente B, Harries E, Marcos JF, Gonzalez-Candelas L, Gabaldon T: **Genome sequence of the necrotrophic fungus *Penicillium digitatum*, the main postharvest pathogen of citrus.** *BMC Genomics* 2012, **13**:646.
46. Liras P, Martin JF: **Gene clusters for beta-lactam antibiotics and control of their expression: why have clusters evolved, and from where did they originate?** *Int Microbiol* 2006, **9**(1):9–19.
47. Martin J, Garcia-Estrada C, Rumbero A, Recio E, Albillos SM, Ullan RV, Martin JF: **Characterization of an Autoinducer of Penicillin Biosynthesis in *Penicillium chrysogenum*.** *Appl Environ Microbiol* 2011, **77**(16):5688–5696.
48. Maskey RP, Grun-Wollny I, Laatsch H: **Isolation, structure elucidation and biological activity of 8-O-methylaverufin and 1,8-O-dimethylaverantin as new antifungal agents from *Penicillium chrysogenum*.** *J Antibiot* 2003, **56**(5):459–463.
49. Ammann AJ: **SMURFs: standard monitoring units for the recruitment of temperate reef fishes.** *J Exp Mar Biol Ecol* 2004, **299**(2):135–154.
50. Marahiel MA: **Working outside the protein-synthesis rules: insights into non-ribosomal peptide synthesis.** *J Pept Sci* 2009, **15**(12):799–807.
51. Walton JD: **HC-toxin.** *Phytochemistry* 2006, **67**(14):1406–1413.
52. Sugang R, Kuo A, Tian XJ, Salerno W, Parikh A, Feasley CL, Dalin E, Tu H, Huang EY, Barry K, Lindquist E, Shapiro H, Bruce D, Schmutz J, Salamov A, Fey P, Gaudet P, Anjard C, Babu MM, Basu S, Bushmanova Y, van der Wel H, Katoh-Kurasawa M, Dinh C, Coutinho PM, Saito T, Elias M, Schaap P, Kay RR, Henrissat B: **Comparative genomics of the social amoebae *Dictyostelium discoideum* and *Dictyostelium purpureum*.** *Genome Biol* 2011, **12**(2):R20.
53. Woo PCY, Lam CW, Tam EWT, Leung CKF, Wong SSS, Lau SKP, Yuen KY: **First Discovery of Two Polyketide Synthase Genes for Mitorubrinic Acid and Mitorubrinol Yellow Pigment Biosynthesis and Implications in Virulence of *Penicillium marneffe*.** *Plos Neglected Trop Dis* 2012, **6**(10):e1871.
54. Woo PCY, Tam EWT, Chong KTK, Cai JJ, Tung ET, Ngan AHY, Lau SKP, Yuen KY: **High diversity of polyketide synthase genes and the melanin biosynthesis gene cluster in *Penicillium marneffe*.** *FEBS J* 2010, **277**(18):3750–3758.
55. Blanc G, Argarkova I, Grimwood J, Kuo A, Brueggeman A, Dunigan DD, Gurnon J, Ladunga I, Lindquist E, Lucas S, Pangilinan J, Proschold T, Salamov A, Schmutz J, Weeks D, Yamada T, Lomsadze A, Borodovsky M, Claverie JM, Grigoriev IV, Van Etten JL: **The genome of the polar eukaryotic microalga *Coccomyxa subellipsoidea* reveals traits of cold adaptation.** *Genome Biol* 2012, **13**(5):R39.
56. Kwan DH, Schulz F: **The Stereochemistry of Complex Polyketide Biosynthesis by Modular Polyketide Synthases.** *Molecules* 2011, **16**(7):6092–6115.
57. Austin MB, Noel AJP: **The chalcone synthase superfamily of type III polyketide synthases.** *Nat Prod Rep* 2003, **20**(1):79–110.
58. Abe I, Morita H: **Structure and function of the chalcone synthase superfamily of plant type III polyketide synthases.** *Nat Prod Rep* 2010, **27**(6):809–838.
59. Goodin S, Kane MP, Rubin EH: **Epithilones: Mechanism of action and biologic activity.** *J Clin Oncol* 2004, **22**(10):2015–2025.
60. Bernhardt R: **Cytochromes P450 as versatile biocatalysts.** *J Biotechnol* 2006, **124**(1):128–145.
61. Guengerich FP: **Cytochrome P450 and chemical toxicology.** *Chem Res Toxicol* 2008, **21**(1):70–83.
62. Chen SL, Xu J, Liu C, Zhu YJ, Nelson DR, Zhou SG, Li CF, Wang LZ, Guo X, Sun YZ, Luo HM, Li Y, Song JY, Henrissat B, Levasseur A, Qian J, Li JQ, Luo X, Shi LC, He L, Xiang L, Xu XL, Niu YY, Li QS, Han MV, Yan HX, Zhang J, Chen HM, Lv AP, Wang Z: **Genome sequence of the model medicinal mushroom *Ganoderma lucidum*.** *Nat Commun* 2012, **3**:913.
63. Strobel T, Al-Dilaimi A, Blom J, Gessner A, Kalinowski J, Luzhetskya M, Puhler A, Szczepanowski R, Bechthold A, Ruckert C: **Complete genome sequence of *Saccharothrix espanaensis* DSM 44229(T) and comparison to the other completely sequenced *Pseudonocardia* species.** *BMC Genomics* 2012, **13**:465.
64. Kikuchi T, Cotton JA, Dalzell JJ, Hasegawa K, Kanzaki N, McVeigh P, Takanashi T, Tsai IJ, Assefa SA, Cock PJA, Otto TD, Hunt M, Reid AJ, Sanchez-Flores A, Tsuchihara K, Yokoi T, Larsson MC, Miwa J, Maule AG, Sasahashi N, Jones JT, Berriman M: **Genomic Insights into the Origin of Parasitism in the Emerging Plant Pathogen *Bursaphelenchus xylophilus*.** *Plos Pathog* 2011, **7**(9):e1002219.
65. Park J, Lee S, Choi J, Ahn K, Park B, Park J, Kang S, Lee YH: **Fungal cytochrome p450 database.** *BMC Genomics* 2008, **9**:402.
66. Moktali V, Park J, Fedorova-Abrams ND, Park B, Choi J, Lee YH, Kang S: **Systematic and searchable classification of cytochrome P450 proteins encoded by fungal and oomycete genomes.** *BMC Genomics* 2012, **13**:525.
67. Deng JX, Carbone I, Dean RA: **The evolutionary history of Cytochrome P450 genes in four filamentous *Ascomycetes*.** *BMC Evol Biol* 2007, **7**:30.
68. van den Brink HJM, van Gorcom RFM, van den Hondel CAMJ, Punt PJ: **Cytochrome P450 enzyme systems in fungi.** *Fungal Genet Biol* 1998, **23**(1):1–17.
69. Hannemann F, Bichet A, Ewen KM, Bernhardt R: **Cytochrome P450 systems - biological variations of electron transport chains.** *BBA-Gen Subjects* 2007, **1770**(3):330–344.
70. Veiga T, Solis-Escalante D, Romagnoli G, ten Pierick A, Hanemaaijer M, Deshmukh A, Wahl A, Pronk JT, Daran JM: **Resolving Phenylalanine Metabolism Sheds Light on Natural Synthesis of Penicillin G in *Penicillium chrysogenum*.** *Eukaryot Cell* 2012, **11**(2):238–249.
71. Luque D, Gonzalez JM, Garriga D, Ghabrial SA, Havens WM, Trus B, Verdaguier N, Carrascosa JL, Caston JR: **The T=1 Capsid Protein of *Penicillium chrysogenum* Virus Is Formed by a Repeated Helix-Rich Core Indicative of Gene Duplication.** *J Virol* 2010, **84**(14):7256–7266.
72. Jiang DH, Ghabrial SA: **Molecular characterization of *Penicillium chrysogenum* virus: reconsideration of the taxonomy of the genus *Chrysovirus*.** *J Gen Virol* 2004, **85**:2111–2121.

73. Gallie DR, Walbot V: **Identification of the Motifs within the Tobacco Mosaic-Virus 5'-Leader Responsible for Enhancing Translation.** *Nucleic Acids Res* 1992, **20**(17):4631–4638.
74. Liu HQ, Fu YP, Jiang DH, Li GQ, Xie JT, Cheng JS, Peng YL, Ghabrial SA, Yi XH: **Widespread Horizontal Gene Transfer from Double-Stranded RNA Viruses to Eukaryotic Nuclear Genomes.** *J Virol* 2010, **84**(22):11876–11887.
75. Bertsch C, Beuve M, Dolja W, Wirth M, Pelsy F, Herrbach E, Lemaire O: **Retention of the virus-derived sequences in the nuclear genome of grapevine as a potential pathway to virus resistance.** *Biol Direct* 2009, **4**:21.
76. Bejarano ER, Khashoggi A, Witty M, Lichtenstein C: **Integration of multiple repeats of geminiviral DNA into the nuclear genome of tobacco during evolution.** *Proc Natl Acad Sci U S A* 1996, **93**(2):759–764.
77. Taylor DJ, Bruenn J: **The evolution of novel fungal genes from non-retroviral RNA viruses.** *BMC Biol* 2009, **7**:88.
78. Felten A, Grandry B, Lagrange PH, Casin I: **Evaluation of three techniques for detection of low-level methicillin-resistant *Staphylococcus aureus* (MRSA): a disk diffusion method with ceftoxitin and moxalactam, the Vitek 2 system, and the MRSA-screen latex agglutination test.** *J Clin Microbiol* 2002, **40**(8):2766–2771.
79. Hosler P, Johnson MJ: **Penicillin from Chemically Defined Media.** *Ind Eng Chem* 1953, **45**(4):871–874.
80. Raeder U, Broda P: **Rapid Preparation of DNA from Filamentous Fungi.** *Lett Appl Microbiol* 1985, **1**(1):17–20.
81. Zerbino DR, Birney E: **Velvet: Algorithms for de novo short read assembly using de Bruijn graphs.** *Genome Res* 2008, **18**(5):821–829.
82. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W: **Scaffolding pre-assembled contigs using SSPACE.** *Bioinformatics* 2011, **27**(4):578–579.
83. Stanke M, Steinkamp R, Waack S, Morgenstern B: **AUGUSTUS: a web server for gene finding in eukaryotes.** *Nucleic Acids Res* 2004, **32**:W309–W312.
84. Jurka J, Klonowski P, Dagman V, Pelton P: **Censor - A program for identification and elimination of repetitive elements from DNA sequences.** *Comput Chem* 1996, **20**(1):119–121.
85. Schattner P, Brooks AN, Lowe TM: **The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs.** *Nucleic Acids Res* 2005, **33**:W686–W689.
86. 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.
87. Darling AE, Mau B, Perna NT: **progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement.** *Plos ONE* 2010, **5**(6):e11147.
88. Krumsiek J, Arnold R, Rattei T: **Gepard: a rapid and sensitive tool for creating dotplots on genome scale.** *Bioinformatics* 2007, **23**(8):1026–1028.
89. Tae H, Kong EB, Park K: **ASMPKS: an analysis system for modular polyketide synthases.** *BMC Bioinformatics* 2007, **8**:327.
90. Finn RD, Mistry J, Tate J, Coghill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, Holm L, Sonnhammer ELL, Eddy SR, Bateman A: **The Pfam protein families database.** *Nucleic Acids Res* 2010, **38**:D211–D222.
91. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: **MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods.** *Mol Biol Evol* 2011, **28**(10):2731–2739.
92. Edgar RC: **MUSCLE: a multiple sequence alignment method with reduced time and space complexity.** *BMC Bioinformatics* 2004, **5**:1–19.
93. Andersen MR, Salazar MP, Schaap PJ, van de Vondervoort PJJ, Culley D, Thykaer J, Frisvad JC, Nielsen KF, Albang R, Albermann K, Berka RM, Braus GH, Braus-Stromeyer SA, Corrochano LM, Dai ZY, van Dijk PWM, Hofmann G, Lasure LL, Magnuson JK, Menke H, Meijer M, Meijer SL, Nielsen JB, Samson RA, Stam H, Tsang A, van den Brink JM, Atkins A, Aerts A, Shapiro H: **Comparative genomics of citric-acid-producing *Aspergillus niger* ATCC 1015 versus enzyme-producing CBS 513.88.** *Genome Res* 2011, **21**(6):885–897.
94. Machida M, Asai K, Sano M, Tanaka T, Kumagai T, Terai G, Kusumoto KI, Arima T, Akita O, Kashiwagi Y, Abe K, Gomi K, Horiuchi H, Kitamoto K, Kobayashi T, Takeuchi M, Denning DW, Galagan JE, Nierman WC, Yu JJ, Archer DB, Bennett JW, Bhatnagar D, Cleveland TE, Fedorova ND, Gotoh O, Horikawa H, Hosoyama A, Ichinomiya M, Igarashi R: **Genome sequencing and analysis of *Aspergillus oryzae*.** *Nature* 2005, **438**(7071):1157–1161.
95. Nierman WC, Pain A, Anderson MJ, Wortman JR, Kim HS, Arroyo J, Berriman M, Abe K, Archer DB, Bermejo C, Bennett J, Bowyer P, Chen D, Collins M, Coulsen R, Davies R, Dyer PS, Farman M, Fedorova N, Fedorova N, Feldblyum TV, Fischer R, Fosker N, Fraser A, Garcia JL, Garcia MJ, Goble A, Goldman GH, Gomi K, Griffith-Jones A: **Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*.** *Nature* 2005, **438**(7071):1151–1156.
96. Futagami T, Mori K, Yamashita A, Wada S, Kajiwara Y, Takashita H, Omori T, Takegawa K, Tashiro K, Kuhara S, Goto M: **Genome Sequence of the White Koji Mold *Aspergillus kawachii* IFO 4308, Used for Brewing the Japanese Distilled Spirit Shochu.** *Eukaryot Cell* 2011, **10**(11):1586–1587.

doi:10.1186/1471-2164-15-144

Cite this article as: Peng et al.: Genomic characteristics and comparative genomics analysis of *Penicillium chrysogenum* KF-25. *BMC Genomics* 2014 **15**:144.

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

