Comparative genomics analysis of *Chryseobacterium* sp. KMC2 reveals metabolic pathways involved in keratinous utilization and natural product biosynthesis

Dingrong Kang[abcd*], Saeed Shoaie[c], Samuel Jacquiod[e], Søren J. Sørensen[a], Rodrigo Ledesma-Amaro[b,*]

[a]Section of Microbiology, Department of Biology, University of Copenhagen, Copenhagen, Denmark

[b]Department of Bioengineering and Imperial College Centre for Synthetic Biology, Imperial College London, London, SW7 2AZ, UK

[c]Centre for Host-Microbiome Interactions, Faculty of Dentistry, Oral & Craniofacial Sciences, King's College London, SE1 9RT, UK

[d]TERRA Research and Teaching Centre, Microbial Processes and Interactions (MiPI), Gembloux Agro-Bio Tech, University of Liège, Gembloux, Belgium

[e]Agroécologie, AgroSup Dijon, INRAE, Univ. Bourgogne, Univ. Bourgogne Franche-Comté, F-21000 Dijon, France

[*]Corresponding authors: Rodrigo Ledesma-Amaro (r.ledesma-amaro@imperial.ac.uk) and Dingrong Kang (dingrong.kang@hotmail.com)

Abstract

Several efforts have been made to valorize keratinous materials, an abundant and renewable resource. Despite these attempts to valorize products generated from keratin hydrolysate, either *via* chemical or microbial conversion, they generally remain with an overall low value. In this study, a promising keratinolytic strain from the genus *Chryseobacterium* (*Chryseobacterium* sp. KMC2) was investigated using comparative genomic tools against publicly available reference genomes to reveal the metabolic potential for biosynthesis of valuable secondary metabolites. Genome and metabolic features of four species were compared, shows different gene numbers but similar functional categories. We successfully mined eleven different secondary metabolite gene clusters of interest from the four genomes, including five common ones shared across all genomes. Among the common metabolites, we identified gene clusters involved in biosynthesis of flexirubin-type pigment, microviridin, and siderophore, all showing remarkable conservation across the four genomes. Unique secondary metabolite gene clusters were also discovered, for example, ladderane from *Chryseobacterium* sp. KMC2. Additionally, this study provides a more comprehensive understanding of the potential metabolic pathways of keratin utilization in *Chryseobacterium* sp. KMC2, with the involvement of amino acid metabolism, TCA cycle, glycolysis/gluconeogenesis, propanoate metabolism, and sulfate reduction. This work uncovers the biosynthesis of secondary metabolite gene clusters from four keratinolytic *Chryseobacterium* spp. and shades lights on the keratinolytic potential of *Chryseobacterium* sp. KMC2 from a genome-mining perspective, providing alternatives to valorize keratinous materials into high-value natural products.

47 **Importance**

48 Keratin is an abundant and renewable resource from slaughterhouses or the poultry industry.
49 Low-value products such as animal feed and fertilizer were generated from these feedstocks
50 based on conventional processing like chemical conversion. In fact, microorganisms possess the
51 potential to synthesize valuable natural products. In this work, we explored the metabolic
52 potential of *Chryseobacterium* sp. KMC2, which was isolated with efficient keratinolytic
53 capacity from a previous study. Comparative genomics analysis displayed similar functional
54 categories against three publicly available reference genomes of keratin-
55 degrading *Chryseobacterium* spp.. Eleven different secondary metabolite gene clusters of interest
56 were mined among four genomes, including five common and unique ones. Furthermore, we
57 provide a more comprehensive understanding of metabolic pathways on keratin utilization
58 in *Chryseobacterium* sp. KMC2, with the involvement of amino acid assimilation and sulfate
59 reduction. These findings contribute to expanding the application of *Chryseobacterium* sp.
60 KMC2 on the valorization of keratinous materials.
61
62 **Keywords:** keratinous materials, metabolic potential, genome mining, gene clusters,
63 degradation pathways
64
65
66
67
68
69

Introduction

Keratin is the most abundant proteins in epithelial cells, constituting the bulk of epidermal appendages such as hair and feather (1, 2). Keratinous materials represent an abundant protein source, particularly originating from the commercial slaughterhouses or poultry farms (3). They contain peptides and amino acids, which are renewable natural resources with great potential in sustainable development (4). However, keratin is an insoluble protein with highly cross-linked disulfide bonds giving it a tough and recalcitrant structure (5). Many attempts have been made to hydrolysis keratinous materials in terms of physicochemical treatment, enzymatic hydrolysis, and microbial conversion (6, 7). The hydrolysis products of keratinous materials have been used for animal feed (8) and fertilizer (6, 9) based on conventional processing.

Microorganisms represent one of the most important natural sources, which have the potential to generate bioactive compounds such as antibiotics, biofuels, and natural pigments derived from cellular metabolites (10, 11). For example, *Yarrowia lipolytica* has been used to convert different renewable feedstocks to high-value metabolites (12). Similarly, *Escherichia coli* has become one of the best cell reactors to produce alcohols, organic acids, biodiesel, even hydrogen by utilizing renewable resources (13). Other bacteria such as *Bacillus subtilis* (14), *Caldicellulosiruptor bescii* (15), *Corynebacterium glutamicum* (16), and *Ruminococcaceae* bacterium (17) were identified and evaluated with the capacity to generate different products by converting renewable carbon sources. Notably, some microorganisms were reported to degrade keratinous waste effectively (18). Exploring keratinolytic potential of these microbes to generate high value-added products is an important step to recycle and valorize keratinous materials.

Molecular mechanisms of microbial keratin degradation are still not fully understood, while genome sequencing offers possibilities to reveal the metabolic potential behind efficient microbial degradation (19). Novel keratinolytic enzymes were identified from the genome of *Bacillus pumilus* 8A6, an efficient keratin degrader (20). Furthermore, going beyond the degradation reaction itself, genomes can also be minded for valuable accessory functions of interest, adding more values to the microbial conversion processes. For instance, gene clusters and biosynthesis pathways of secondary metabolites could be disclosed from genomes *via* adequate analysis tools (21, 22). A total of 104 putative biosynthetic gene clusters (BGCs) for secondary metabolites were predicted from nine *Ktedonobacteria* genomes (23). Secondary metabolites were identified and linked to gene clusters based on the comparison and mining of six genomes belonging to diverse *Aspergillus* species, successfully fueling industrial biotechnology initiatives and medical research (24). Therefore, using the genomes of keratinolytic microbial species in a similar way would represent a promising approach to discover biosynthetic gene clusters of secondary metabolites of interest, excavating the full application potential of these microbes.

Recently, several studies based on different environments have revealed the remarkable potential of representative taxa from the *Chryseobacterium* genus for keratin degradation using isolation, activity tests and genome sequencing (25, 26). In this study, a novel strain *Chryseobacterium* sp. KMC2 was previously obtained from an enrichment procedure, displaying a potent capacity of keratin degradation (27). The genome of *Chryseobacterium* sp. KMC2 was analyzed and compared with publicly available genomes of other keratinolytic *Chryseobacterium* spp. to clarify the genomic basis of keratin degradation, and to unravel hidden biosynthetic gene clusters

of interest. Subsequently, the metabolic pathways associated with keratin degradation were constructed, providing deeper insight into the yet obscure keratinolytic processes. This work reveals the keratinolytic potential of *Chryseobacterium* spp. and mined potential accessory gene clusters of secondary metabolites, which could i) contribute to optimizing the processes of keratin degradation and ii) broaden the perspective to generate added-value products from keratin hydrolysate.

Results and discussion

Genome feature comparison of four keratinolytic *Chryseobacterium* strains

Chryseobacterium sp. KMC2 originated from a river-bank soil sample, and displayed a potent degradation ability toward milled pig bristle and hooves (27, 28). The genome of *Chryseobacterium* sp. KMC2 was sequenced and compared to three reference genomes of *Chryseobacterium* spp. Including: i) *Chryseobacterium camelliae* Dolsongi-HT, isolated from green tea leaves (29); ii) *Chryseobacterium gallinarum* strain DSM 27622, isolated from chicken (30); and iii) *Chryseobacterium* sp. strain P1-3 isolated from poultry waste (31), which all display keratinolytic capacity (Table 1). *Chryseobacterium* sp. KMC2 showed distinct genome feature from the other known keratinolytic strains. The genome size of *Chryseobacterium* sp. KMC2 is 5.28 Mbp, larger than the other three genomes, which ranged from 4.38 Mbp to 4.63 Mbp. A total of 4,773 genes were predicted from *Chryseobacterium* sp. KMC2 genome, and about 4,000 genes were annotated from the other three genomes. Besides, the GC content ranges from 36.33% to 41.80% in *Chryseobacterium* spp. genomes. Furthermore, the whole-genome

phylogenetic tree was constructed with other eight publically available *Chryseobacterium* spp. genomes (Fig. 1a), showing the highly similarity among *Chryseobacterium gallinarum* strain DSM 27622 and *Chryseobacterium* sp. strain P1-3. Notably, *Chryseobacterium* sp. KMC2 and *Chryseobacterium camelliae* Dolsongi-HT have closer evolutionary relationship with other *Chryseobacterium* species. Although the keratinolytic capacity within these eight strains is still not clear, more keratinolytic species are expected to be found in this genus according to the phylogenic tree analysis. Four strains among this genus have documented with actual keratin degradation capability (29-31), while *Chryseobacterium* sp. KMC2 may possess different keratinolytic potential encoded in its larger genome compared to three other *Chryseobacterium* strains.

Metabolic potential comparison of four keratinolytic *Chryseobacterium* genomes

About 40% of the genes from the four genomes were classified into various functional categories based on the KEGG database. The vast majority of annotated genes belonged to metabolism, genetic information processing, environmental information processing, and cellular processes (Fig. 1b). The functional categories of the genomes were overall highly similar, with ~85% of annotated genes assigned to “metabolism” (category A) which included ~1.000 genes into the sub-category “global and overview maps”. Additionally, about 8% and 4% annotated genes from each genome were assigned to “genetic information processing” (category B) and “environmental information processing” (category C), respectively. The remaining annotated genes belonged to “cellular processes” (category D), which occupied 3% of the annotated genomes approximately.

Remarkably, each genome had more than 200 genes assigned into the "amino acid metabolism" sub-category. Keratin is mainly composed of amino acids (3), which is ultimately the operational carbon nutrient source exploited for microbial growth. Numerous amino acid metabolism-related enzymes were annotated, revealing the genetic potential of these *Chryseobacterium* strains for using keratin materials as energy sources. Of particular interest, several biosynthesis genes of secondary metabolites were detected from the genomes, of which more than 20 genes were annotated as “metabolism of terpenoids and polyketides” and around 40 genes were annotated as “biosynthesis of other secondary metabolites” sub-category (Fig. 1b). Terpenoids are a group of natural products with diverse commercial applications, which have been produced from microbial cell factories (32). Many polyketides are considered as significant natural products with broad applications in the agriculture and pharmaceutical industry (33). The metabolic pathways related to polyketides biosynthesis production are well understood in some microorganisms like *Streptomyces* which play a crucial role in industrial bioproduction (34). This result indicates that these *Chryseobacterium* strains could have the potential to synthesize high-value secondary metabolites such as terpenoids and polyketides from keratinous materials.

Mining and comparing secondary metabolite gene clusters

Genome mining is an effective approach to discover new natural products from microorganisms based on “signature genes” detection or searching for specific patterns in gene sequences (35). To explore the potential of producing high value chemicals from these four *Chryseobacterium*, secondary metabolite gene clusters were predicted by using antiSMASH 5.0 mining pipeline (Fig. 2). In total, eleven different secondary metabolite gene clusters were identified. *Chryseobacterium* sp. KMC2 possesses the largest number (15), while *Chryseobacterium*

185 *camelliae* Dolsongi-HT1 has the fewest (8). Ten gene clusters were predicted from the other two
186 strains. Five clusters are present in the four genomes, which are flexirubin-type pigment
187 (resorcinol and arylpolyene), microviridin, lanthipeptide, NRPS-like, and siderophore.
188 Remarkably, the flexirubin-type pigment is a typical metabolite produced from *Flavobacterium*
189 (36). Several species from *Chryseobacterium* were previously designated and known
190 as *Flavobacterium* owes to similar characteristics including the presence of yellow pigments (37).
191 Flexirubin-type pigment was isolated and characterized from *Chryseobacterium* sp. UTM-3T
192 (38). In addition, *Chryseobacterium* sp. KMC2 owns a unique gene cluster to produce ladderane.
193 Another unique natural product is beta-lactone from *Chryseobacterium camelliae* Dolsongi-HT1.
194 Ladderanes are hydrocarbon chains which were regarded as membrane lipid components
195 produced by anammox (anaerobic ammonia-oxidizing) bacteria uniquely, but the production is
196 not affordable due to their extremely low growth (39, 40). These results demonstrate that various
197 secondary metabolite gene clusters including both expected and unusual candidates were
198 discovered from *Chryseobacterium* genomes, which could turn into novel natural product
199 sources.
200
201 **Synteny analysis and features of secondary metabolite gene clusters**
202 Comparative genomics can reveal unique cluster and distribution patterns of secondary
203 metabolites in species (41). Five secondary metabolite gene clusters, predicted to be present in
204 the four genomes, were selected to explore the evolutionary relationship among four
205 *Chryseobacterium* strains. Three of them including flexirubin-type pigment, microviridin, and
206 siderophore display a conserved gene cluster structure from synteny analysis (Fig. 3, 4, and 5),
207 while the other two showed no evident synteny relation (Fig. S1 and S2).

Flexirubin-type pigment. Natural pigments have increasing applications in food, pharmaceutical, and textile industries, owing to their advantages such as non-toxic, biodegradable, and low allergenic potential compared to synthetic pigments (42). In particular, flexirubin-type pigment has a potential antimicrobial and anti-tumoral activities (43). Biosynthesis gene clusters of flexirubin-type pigment are conserved across the four tested genomes, especially within *Chryseobacterium gallinarum* strain DSM 27622 and *Chryseobacterium* sp. strain P1-3 (Fig. 3a). A total of 61 biosynthesis-related genes of flexirubin-type pigment were predicted from *Chryseobacterium* sp. KMC2, including four core biosynthesis genes. One of the core biosynthesis genes was annotated as 3-oxoacyl-(acyl carrier protein) synthase III (*Flex11*), and the other three were annotated as Beta-ketoacyl synthases (*Flex21*, *Flex24*, and *Flex40*). Besides, transport-related genes and regulatory genes were predicted from the gene cluster. A previous study identified the molecular structure of flexirubin-type pigment isolated from a *Chryseobacterium* sp. UTM-3T (38). According to the products from core biosynthesis genes and their molecular structures, a proposition of biosynthesis pathway was established (Fig. 3b), where flexirubin-type pigment is generated from resorcinol and arylpolyene. Further transcriptomics and metabolomics analysis would be required to confirm the validity of this potential pathway discovery.

Microviridin. Microviridins represent a group of ribosomally synthesized peptides under post-translational modifications, which have been mainly isolated from cyanobacteria and present potent serine-type protease inhibitory activities (44, 45). These properties could make microviridin serve as the natural antimicrobial agents for developing potential drugs.

231 Biosynthesis gene clusters of microviridin from four *Chryseobacterium* genomes show a highly
232 conserved structure with a similarity greater than 71% from most gene synteny analysis (Fig. 4a).
233 There are 23 biosynthesis genes of microviridin that were predicted from *Chryseobacterium* sp.
234 KMC2. Two core biosynthetic genes (A and B) were identified from genomes and transport-
235 related genes were also been discovered. Besides, amino acid sequences of mvdA and mvdB
236 were aligned, showing that multiple motifs from mvdA and mvdB are conserved (Fig. 4bc).
237 Interestingly, many keratinases were reported to be classified as serine proteases, acting on the
238 molecular structure of keratin (46). This suggests that microviridins may regulate keratinolytic
239 activity, further characterizing and manipulating the microviridin synthetic pathway could
240 contribute to improving the keratin degradation efficiency.
241
242 **Siderophore.** Siderophores are ferric ion-specific chelators to scavenge iron from the
243 extracellular environment, which play important roles in virulence and oxidative stress tolerance
244 in microorganisms (47). It has been designed as a Trojan horse antibiotic to enter and kill
245 pathogenic bacteria (48), and also has shown the potential to decrease the growth of cancerous
246 cells (49). Biosynthesis gene cluster of siderophore shows a high synteny conservation among
247 *Chryseobacterium* sp. KMC2 and *Chryseobacterium camelliae* Dolsongi-HT1,
248 *Chryseobacterium gallinarum* strain DSM 27622 and *Chryseobacterium* sp. Strain P1-3,
249 respectively (Fig. 5). A total of ten genes were predicted from siderophore biosynthesis cluster of
250 *Chryseobacterium* sp. KMC2, and eight genes from the other three *Chryseobacterium* strains
251 separately. Functional description of each gene related to siderophore biosynthesis in
252 *Chryseobacterium* sp. KMC2 shows two core biosynthesis genes, and includes one regulatory
253 gene and one transport-related gene. This further suggests that those siderophore are potentially

254 fully functional molecular features that can be regulated on-demand and exported outside the cell
255 when needed.
256
257 **Metabolic pathways of keratin utilization in *Chryseobacterium* sp. KMC2 genome**
258 The main metabolic pathways related to keratin utilization in *Chryseobacterium* sp. KMC2
259 genome were investigated. These pathways included amino acid metabolism, TCA cycle,
260 glycolysis/gluconeogenesis, propanoate metabolism, and sulfate reduction (Fig. 6). A previous
261 study suggested that abundant amino acids are released during microbial degradation and used as
262 nutrient sources, such as leucine and aspartate (28). The metabolic pathways of amino acid
263 utilization were mapped from the genome of *Chryseobacterium* sp. KMC2. Most of the amino
264 acids are converted into intermediates of the TCA cycle. For instance, arginine can be converted
265 to succinate, then enter to TCA cycle after a multiple-steps enzyme reaction. Aspartate, tyrosine,
266 phenylalanine, and glutamate could serve as the substrates to generate fumarate, thus being part
267 of the TCA cycle. Besides, isoleucine turns into the substrates of 2-methyl-acetoacetyl-CoA after
268 several enzymatic steps, which is then converted into acetyl-CoA and propanoyl-CoA *via* acetyl-
269 CoA C-acyltransferase. Acetyl-CoA is an important intermediate, which can entry to the TCA
270 cycle *via* citrate synthase (50). It is also the precursors of fatty acid and polyketides biosynthesis
271 (51). Propanoyl-CoA serves as the critical substrate within propanoate metabolism and can also
272 be used to make lipids (52, 53). On the other hand, methionine can be converted to 2-
273 oxobutanoate, which is also an intermediate of propanoate metabolism. Subsequently, the
274 methylmalonyl-CoA generated in propanoate metabolism enters into the TCA cycle *via* succinyl-
275 CoA. Besides, the key enzymes of glycolysis/gluconeogenesis were found, indicating the
276 potential to produce essential biomass components based on oxaloacetate from TCA cycle.

Evidence indicates that a source of redox is needed for complete keratin degradation with keratinases (54, 55). Several metabolites such as sulfite were revealed to be associated with efficient keratin degradation (56). Therefore, here the complete metabolic pathway of sulfate reduction was mapped in the genome, which shows that the potential to create a redox environment needed for keratinases is indeed present. Following the development of sequencing technologies, increasing genomes of keratinolytic species have been unveiled, which provide a genomic perspective to reveal the molecular keratinolytic mechanisms. For instance, metabolic pathways related to keratin degradation such as enzymolysis and reduction of disulfide bonds were clarified through uncovering the genetic basis of microbial genomes (57). The complex keratinolytic processes of *Streptomyces* sp. included protease secretion, iron uptake, spore formation, and resuscitation were recently revealed from a genome view (58). Our results are in line with the notion that a redox environment is indeed required for efficient keratinolytic activity to occur. It is expected that the integrated metabolic pathways associated with keratinolytic processes will be deciphered along with more genomes sequencing and biochemical studies of relevant metabolic pathways.

Conclusion

In this work, the genomes from four *Chryseobacterium* spp. with keratinolytic activity were analyzed. Common and unique secondary metabolite gene clusters were mined from *Chryseobacterium* spp. genomes, suggesting the potential to generate high value metabolites using keratin-rich wastes as the nutrient sources. Therefore, the use of these microorganisms could be an alternative way to valorize keratinous materials through microbial conversion.

Furthermore, the metabolic pathways of keratin degradation from *Chryseobacterium* sp. KMC2 was studied from a genomic viewpoint. Nevertheless, there are still unknowns to link both metabolic pathways of keratinous utilization and the natural products biosynthesis. Understanding these connected pathways and their regulation will contribute to developing synthetic biology approaches to boost high value-added products from microbial keratin degradation.

Materials and Methods

DNA preparation

Chryseobacterium sp. KMC2 was isolated and identified from a keratinolytic microbial consortium enriched from a soil sample (19, 27, 28). The keratinolytic capacity of *Chryseobacterium* sp. KMC2 was evaluated as the keystone strain among the microbial consortium, which was able to degrade keratin materials efficiently (27). *Chryseobacterium* sp. KMC2 was inoculated to LB medium, and cultured overnight (200 rpm, 30 °C). Two milliliters of the suspension were centrifuged and collected to prepare the DNA extraction, performed by using by FAST Soil DNA Kit (MP Biomedicals, United States) according to the manufacturer's instructions.

Genome sequencing, assembling, and functional annotation

The genome sequencing was performed by an Illumina Miseq instrument at the University of Copenhagen by using DNA Library Preparation Kits v2 (2 × 250 bp), according to the manufacturer's instructions. Raw reads were treated and assembled to contigs on CLC Genomic

Workbench 8.5.1. The obtained contigs were validated using QUAST 4.5 (59). Genes were predicted from the contigs and further annotated with Prokka v1.14.5 (60). Predicted genes from contigs were submitted to eggNOG 5.0 database to obtain an integrated functional annotation and classification (61).

Whole-genome phylogenetic analysis

To determine the phylogenetic origin of *Chryseobacterium* sp. KMC2 in the *Chryseobacterium* genus, the whole-genome sequences of 11 publicly available *Chryseobacterium* spp. were downloaded from NCBI database to construct a phylogenetic tree. The whole-genome sequence-based phylogenetic tree was inferred by using an online pipeline: The Reference sequence Alignment based Phylogeny builder (REALPHY 1.12) (62), based on the merge reference alignments. Then the visualization of the phylogenetic tree was generated by iTOL v5 (63).

Secondary metabolite gene cluster detection and annotation

Assembled contigs of four *Chryseobacterium* spp. were uploaded to antiSMASH 5.0 secondary metabolite genome mining web platform (21). Predicted secondary metabolites gene clusters from *Chryseobacterium* sp. KMC2 were compared with other keratinolytic *Chryseobacterium* strains. Gene annotation of each cluster from *Chryseobacterium* sp. KMC2 was performed by Prokka v1.14.5 (60) and BLASTP with the NCBI database. The best match sequencing ID was recorded for the annotated genes. Synteny and features of conservative secondary metabolite gene clusters were analyzed by using Easyfig 2.2.2, showing the similarity of gene sequences (64). Feature comparison of amino acid sequences and motifs from core synthetic genes were

345 analyzed by using Clustal Omega (65) to get the multiple sequence alignment and using
346 Seq2logo to generate sequence logo (66).
347
348 **Metabolic networks construction**
349 The genomes of *Chryseobacterium* sp. KMC2 and other three *Chryseobacterium* spp. were
350 submitted to GhostKOALA (67) to obtain the KO number for each gene, then genes were
351 assigned to different metabolic pathways and functional categories. Following the metabolic
352 networks construction of *Chryseobacterium* sp. KMC2 was achieved through mapping the
353 annotated enzyme genes to KEGG (68) reference pathway and Biocyc database (69) manually.
354
355 **Data availability**
356 Reference genomes were downloaded from the NCBI database: *Chryseobacterium camelliae*
357 strain Dolsongi-HT1 (GenBank: GCA_002770595.1), *Chryseobacterium gallinarum* strain DSM
358 27622 (GenBank: GCA_001021975.1), *Chryseobacterium* sp. P1-3 (GenBank:
359 GCA_000738495.1), *Chryseobacterium gleum* NCTC11432 (GenBank: GCA_900636535.1),
360 *Chryseobacterium bernardetii* H4638 (GenBank: GCA_003815955.1), *Chryseobacterium*
361 arthrosphaerae FDAAGOS 519 (GenBank: GCA_003812705.1), *Chryseobacterium indologenes*
362 FDAARGOS 337 (GenBank: GCA_002208925.2), *Chryseobacterium joostei* DSM 16927
363 (GenBank: GCA_003815775.1), *Chryseobacterium glaciei* IHBB 10212 (GenBank:
364 GCA_001648155.1), *Chryseobacterium carnipullorum* F9942 (GenBank: GCA_003815855.1),
365 and *Chryseobacterium* sp. SNU WT5 (GenBank: GCA_007362475.1). Raw sequencing data
366 were deposited in the Sequence Read Archive (SRA) database under the BioProject number
367 PRJNA686768 with an accession number SRR13278108. The assembled genome sequence

368 of *Chryseobacterium* sp. KMC2 has been deposited at DDBJ/ENA/GenBank under the accession
369 JAESIT000000000.
370
371 **Acknowledgements**
372 This research was funded by the Danish Innovation Fund (Grant Number 1308-00015B,
373 Keratin2Protein) and also under the support of the Chinese Scholarship Council (CSC)
374 Scholarship Program.
375
376
377 **References**
378 1. Wang B, Yang W, McKittrick J, Meyers MA. 2016. Keratin: Structure, mechanical
379 properties, occurrence in biological organisms, and efforts at bioinspiration. Progress in
380 Materials Science 76:229-318.
381 2. Coulombe PA, Omary MB. 2002. 'Hard'and 'soft'principles defining the structure,
382 function and regulation of keratin intermediate filaments. Current opinion in cell biology 14:110-
383 122.
384 3. Korniłłowicz-Kowalska T, Bohacz J. 2011. Biodegradation of keratin waste: theory and
385 practical aspects. Waste management 31:1689-1701.
386 4. Khosa M, Ullah A. 2013. A sustainable role of keratin biopolymer in green chemistry: a
387 review. J Food Processing & Beverages 1:8.
388 5. Rajabi M, Ali A, McConnell M, Cabral J. 2020. Keratinous materials: Structures and
389 functions in biomedical applications. Materials Science and Engineering: C 110:110612.

6. Holkar CR, Jain SS, Jadhav AJ, Pinjari DV. 2018. Valorization of keratin based waste. Process Safety and Environmental Protection 115:85-98.

7. Zhao W, Yang R, Zhang Y, Wu L. 2012. Sustainable and practical utilization of feather keratin by an innovative physicochemical pretreatment: high density steam flash-explosion. Green chemistry 14:3352-3360.

8. Dios D. 2001. Fishmeal replacement with feather-enzymatic hydrolyzates co-extruded with soya-bean meal in practical diets for the Pacific white shrimp (*Litopenaeus vannamei*). Aquaculture Nutrition 7:143-151.

9. Gaidau C, Epure D-G, Enascuta CE, Carsote C, Sendrea C, Proietti N, Chen W, Gu H. 2019. Wool keratin total solubilisation for recovery and reintegration-An ecological approach. Journal of Cleaner Production 236:117586.

10. Donadio S, Monciardini P, Alduina R, Mazza P, Chiocchini C, Cavaletti L, Sosio M, Puglia AM. 2002. Microbial technologies for the discovery of novel bioactive metabolites. Journal of Biotechnology 99:187-198.

11. Smitha M, Singh S, Singh R. 2017. Microbial biotransformation: a process for chemical alterations. J Bacteriol Mycol Open Access 4:85.

12. Ledesma-Amaro R, Nicaud J-M. 2016. Metabolic engineering for expanding the substrate range of *Yarrowia lipolytica*. Trends in Biotechnology 34:798-809.

13. Zhao C, Zhang Y, Li Y. 2019. Production of fuels and chemicals from renewable resources using engineered *Escherichia coli*. Biotechnology advances 37:107402.

14. Chen C, Lin J, Wang W, Huang H, Li S. 2019. Cost-effective production of surfactin from xylose-rich corncob hydrolysate using *Bacillus subtilis* BS-37. Waste and Biomass Valorization 10:341-347.

413 15. Straub CT, Bing RG, Wang JP, Chiang VL, Adams MW, Kelly RM. 2020. Use of the
414 lignocellulose-degrading bacterium *Caldicellulosiruptor bescii* to assess recalcitrance and
415 conversion of wild-type and transgenic poplar. Biotechnology for biofuels 13:1-10.
416 16. Sgobba E, Blöbaum L, Wendisch VF. 2018. Production of food and feed additives from
417 non-food-competing feedstocks: Valorizing N-acetylmuramic acid for amino acid and carotenoid
418 fermentation with *Corynebacterium glutamicum*. Frontiers in Microbiology 9:2046.
419 17. Zhu X, Zhou Y, Wang Y, Wu T, Li X, Li D, Tao Y. 2017. Production of high-
420 concentration n-caproic acid from lactate through fermentation using a newly isolated
421 *Ruminococcaceae* bacterium CPB6. Biotechnology for biofuels 10:102.
422 18. Li Q. 2019. Progress in microbial degradation of feather waste. Frontiers in Microbiology
423 10:2717.
424 19. Kang D, Huang Y, Nesme J, Herschend J, Jacquiod S, Kot W, Hansen LH, Lange L,
425 Sørensen SJ. 2020. Metagenomic analysis of a keratin-degrading bacterial consortium provides
426 insight into the keratinolytic mechanisms. Science of The Total Environment:143281.
427 20. Huang Y, Łężyk M, Herbst F-A, Busk PK, Lange L. 2020. Novel keratinolytic enzymes,
428 discovered from a talented and efficient bacterial keratin degrader. Scientific reports 10:1-11.
429 21. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T. 2019.
430 antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic acids
431 research 47:W81-W87.
432 22. Weber T, Kim HU. 2016. The secondary metabolite bioinformatics portal: Computational
433 tools to facilitate synthetic biology of secondary metabolite production. Synthetic and Systems
434 Biotechnology 1:69-79.

435 23. Zheng Y, Saitou A, Wang C-M, Toyoda A, Minakuchi Y, Sekiguchi Y, Ueda K, Takano
436 H, Sakai Y, Abe K. 2019. Genome features and secondary metabolites biosynthetic potential of
437 the class *Ktedonobacteria*. Frontiers in microbiology 10:893.
438 24. Kjærbølling I, Vesth TC, Frisvad JC, Nybo JL, Theobald S, Kuo A, Bowyer P, Matsuda
439 Y, Mondo S, Lyhne EK. 2018. Linking secondary metabolites to gene clusters through genome
440 sequencing of six diverse *Aspergillus* species. Proceedings of the National Academy of Sciences
441 115:E753-E761.
442 25. Fontoura R, Daroit DJ, Corrêa APF, Moresco KS, Santi L, Beys-da-Silva WO, Yates III
443 JR, Moreira JCF, Brandelli A. 2019. Characterization of a novel antioxidant peptide from feather
444 keratin hydrolysates. New biotechnology 49:71-76.
445 26. Kshetri P, Roy SS, Sharma SK, Singh TS, Ansari MA, Prakash N, Ngachan S. 2019.
446 Transforming chicken feather waste into feather protein hydrolysate using a newly isolated
447 multifaceted keratinolytic bacterium *Chryseobacterium sediminis* RCM-SSR-7. Waste and
448 Biomass Valorization 10:1-11.
449 27. Kang D, Jacquiod S, Herschend J, Wei S, Nesme J, Sørensen SJ. 2020. Construction of
450 simplified microbial consortia to degrade recalcitrant materials based on enrichment and
451 dilution-to-extinction cultures. Frontiers in microbiology 10:3010.
452 28. Kang D, Herschend J, Al-Soud WA, Mortensen MS, Gonzalo M, Jacquiod S, Sørensen
453 SJ. 2018. Enrichment and characterization of an environmental microbial consortium displaying
454 efficient keratinolytic activity. Bioresource technology 270:303-310.
455 29. Kim E-M, Hwang KH, Park J-S. 2018. Complete Genome Sequence of
456 *Chryseobacterium camelliae* Dolsongi-HT1, a Green Tea Isolate with Keratinolytic Activity.
457 Genome Announcements 6.

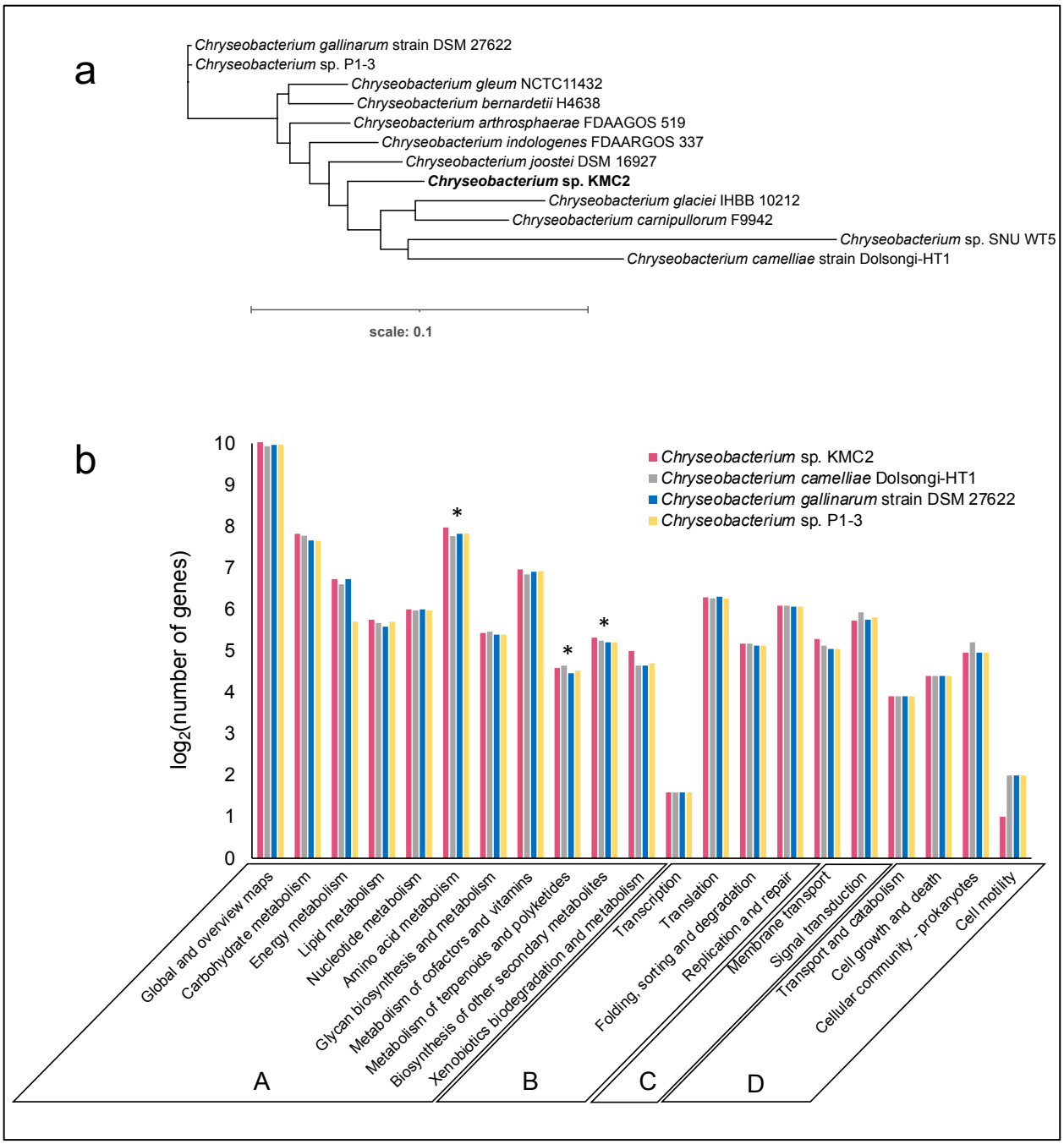
458 30. Park G-S, Hong S-J, Jung BK, Khan AR, Park Y-J, Park CE, Lee A, Kwak Y, Lee Y-J,
459 Lee D-W. 2015. Complete genome sequence of a keratin-degrading bacterium *Chryseobacterium*
460 *gallinarum* strain DSM 27622T isolated from chicken. Journal of biotechnology 211:66-67.
461 31. Park G-S, Hong S-J, Lee C-H, Khan AR, Ullah I, Jung BK, Choi J, Kwak Y, Back C-G,
462 Jung H-Y. 2014. Draft genome sequence of *Chryseobacterium* sp. strain P1-3, a keratinolytic
463 bacterium isolated from poultry waste. Genome announcements 2.
464 32. Lin P-C, Pakrasi HB. 2019. Engineering cyanobacteria for production of terpenoids.
465 Planta 249:145-154.
466 33. Barajas JF, Blake-Hedges JM, Bailey CB, Curran S, Keasling JD. 2017. Engineered
467 polyketides: synergy between protein and host level engineering. Synthetic and systems
468 biotechnology 2:147-166.
469 34. Wang W, Li S, Li Z, Zhang J, Fan K, Tan G, Ai G, Lam SM, Shui G, Yang Z. 2020.
470 Harnessing the intracellular triacylglycerols for titer improvement of polyketides in *Streptomyces*.
471 Nature Biotechnology 38:76-83.
472 35. Adamek M, Spohn M, Stegmann E, Ziemert N. 2017. Mining bacterial genomes for
473 secondary metabolite gene clusters, p 23-47, Antibiotics. Springer.
474 36. Reichenbach H, Kohl W, Böttger-Vetter A, Achenbach H. 1980. Flexirubin-type
475 pigments in *Flavobacterium*. Archives of Microbiology 126:291-293.
476 37. Vandamme P, Bernardet J-F, Segers P, Kersters K, Holmes B. 1994. New Perspectives in
477 the Classification of the Flavobacteria: Description of *Chryseobacterium* gen. nov., *Bergeyella*
478 gen. nov., and *Empedobacter* nom. rev. International Journal of Systematic and Evolutionary
479 Microbiology 44:827-831.

480 38. Venil CK, Zakaria ZA, Usha R, Ahmad WA. 2014. Isolation and characterization of
481 flexirubin type pigment from *Chryseobacterium* sp. UTM-3T. Biocatalysis and Agricultural
482 Biotechnology 3:103-107.
483 39. Javidpour P, Deutsch S, Mutalik VK, Hillson NJ, Petzold CJ, Keasling JD, Beller HR.
484 2016. Investigation of proposed ladderane biosynthetic genes from anammox bacteria by
485 heterologous expression in *E. coli*. PloS one 11:e0151087.
486 40. Mercer JA, Cohen CM, Shuken SR, Wagner AM, Smith MW, Moss III FR, Smith MD,
487 Vahala R, Gonzalez-Martinez A, Boxer SG. 2016. Chemical synthesis and self-assembly of a
488 ladderane phospholipid. Journal of the American Chemical Society 138:15845-15848.
489 41. Adamek M, Alanjary M, Sales-Ortells H, Goodfellow M, Bull AT, Winkler A, Wibberg
490 D, Kalinowski J, Ziemert N. 2018. Comparative genomics reveals phylogenetic distribution
491 patterns of secondary metabolites in *Amycolatopsis* species. BMC genomics 19:1-15.
492 42. Venil CK, Zakaria ZA, Ahmad WA. 2013. Bacterial pigments and their applications.
493 Process Biochemistry 48:1065-1079.
494 43. Aruldass CA, Dufossé L, Ahmad WA. 2018. Current perspective of yellowish-orange
495 pigments from microorganisms-a review. Journal of Cleaner Production 180:168-182.
496 44. Philmus B, Christiansen G, Yoshida WY, Hemscheidt TK. 2008. Post-translational
497 modification in microviridin biosynthesis. ChemBioChem 9:3066-3073.
498 45. Gatte-Picchi D, Weiz A, Ishida K, Hertweck C, Dittmann E. 2014. Functional analysis of
499 environmental DNA-derived microviridins provides new insights into the diversity of the
500 tricyclic peptide family. Applied and environmental microbiology 80:1380-1387.

501 46. Hassan MA, Abol-Fotouh D, Omer AM, Tamer TM, Abbas E. 2020. Comprehensive
502 insights into microbial keratinases and their implication in various biotechnological and
503 industrial sectors: A review. International Journal of Biological Macromolecules.
504 47. Khan A, Singh P, Srivastava A. 2018. Synthesis, nature and utility of universal iron
505 chelator–Siderophore: A review. Microbiological research 212:103-111.
506 48. Boda SK, Pandit S, Garai A, Pal D, Basu B. 2016. Bacterial siderophore mimicking iron
507 complexes as DNA targeting antimicrobials. RSC advances 6:39245-39260.
508 49. Nakouti I, Sihanonth P, Palaga T, Hobbs G. 2013. Effect of a siderophore producer on
509 animal cell apoptosis: a possible role as anti-cancer agent. International Journal of Pharma
510 Medicine and Biological Sciences 2:1-5.
511 50. REMINGTON SJ. 1992. Structure and mechanism of citrate synthase, p 209-229,
512 Current topics in cellular regulation, vol 33. Elsevier.
513 51. Cronan JE, Thomas J. 2009. Bacterial fatty acid synthesis and its relationships with
514 polyketide synthetic pathways. Methods in enzymology 459:395-433.
515 52. Park Y-K, Dulermo T, Ledesma-Amaro R, Nicaud J-M. 2018. Optimization of odd chain
516 fatty acid production by *Yarrowia lipolytica*. Biotechnology for biofuels 11:1-12.
517 53. Park Y-k, Ledesma-Amaro R, Nicaud J-M. 2020. De novo biosynthesis of odd-chain
518 fatty acids in *Yarrowia lipolytica* enabled by modular pathway engineering. Frontiers in
519 Bioengineering and Biotechnology 7:484.
520 54. Nasipuri P, Herschend J, Brejnrod AD, Madsen JS, Espersen R, Svensson B, Burmølle M,
521 Jacquiod S, Sørensen SJ. 2020. Community-intrinsic properties enhance keratin degradation
522 from bacterial consortia. PloS one 15:e0228108.

523 55. Ramnani P, Gupta R. 2007. Keratinases vis-à-vis conventional proteases and feather
524 degradation. World Journal of Microbiology and Biotechnology 23:1537-1540.
525 56. Navone L, Speight R. 2018. Understanding the dynamics of keratin weakening and
526 hydrolysis by proteases. PLoS One 13:e0202608.
527 57. Huang Y, Busk PK, Herbst F-A, Lange L. 2015. Genome and secretome analyses provide
528 insights into keratin decomposition by novel proteases from the non-pathogenic fungus *Onygena*
529 *corvina*. Applied microbiology and biotechnology 99:9635-9649.
530 58. Li Z-W, Liang S, Ke Y, Deng J-J, Zhang M-S, Lu D-L, Li J-Z, Luo X-C. 2020. The
531 feather degradation mechanisms of a new *Streptomyces* sp. isolate SCUT-3. Communications
532 Biology 3:1-13.
533 59. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for
534 genome assemblies. Bioinformatics 29:1072-1075.
535 60. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics
536 30:2068-2069.
537 61. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H,
538 Mende DR, Letunic I, Rattei T, Jensen LJ. 2019. eggNOG 5.0: a hierarchical, functionally and
539 phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses.
540 Nucleic acids research 47:D309-D314.
541 62. Bertels F, Silander OK, Pachkov M, Rainey PB, van Nimwegen E. 2014. Automated
542 reconstruction of whole-genome phylogenies from short-sequence reads. Molecular biology and
543 evolution 31:1077-1088.
544 63. Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new
545 developments. Nucleic acids research 47:W256-W259.

546 64. Sullivan MJ, Petty NK, Beatson SA. 2011. Easyfig: a genome comparison visualizer.
547 Bioinformatics 27:1009-1010.
548 65. Sievers F, Higgins DG. 2018. Clustal Omega for making accurate alignments of many
549 protein sequences. Protein Science 27:135-145.
550 66. Thomsen MCF, Nielsen M. 2012. Seq2Logo: a method for construction and visualization
551 of amino acid binding motifs and sequence profiles including sequence weighting, pseudo counts
552 and two-sided representation of amino acid enrichment and depletion. Nucleic acids research
553 40:W281-W287.
554 67. Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG tools
555 for functional characterization of genome and metagenome sequences. Journal of molecular
556 biology 428:726-731.
557 68. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima
558 S, Okuda S, Tokimatsu T. 2007. KEGG for linking genomes to life and the environment. Nucleic
559 acids research 36:D480-D484.
560 69. Caspi R, Altman T, Billington R, Dreher K, Foerster H, Fulcher CA, Holland TA,
561 Keseler IM, Kothari A, Kubo A. 2014. The MetaCyc database of metabolic pathways and
562 enzymes and the BioCyc collection of Pathway/Genome Databases. Nucleic acids research
563 42:D459-D471.
564
565
566
567
568



569

570 **Fig. 1. Analysis of the *Chryseobacterium* spp. genomes.** (a). The whole-genome sequence-

571 based phylogenetic tree of *Chryseobacterium* spp., constructed by REALPHY 1.12 (62), based

572 on the merge reference alignments of all genomes. Branch length represents divergence. (b).

573 Comparison of KEGG function classification amongst four *Chryseobacterium* spp. genomes.

574 Functional categories: Metabolism (A), Genetic information processing (B), Environmental

575 information processing (C), and Cellular processes (D). The stars show the sub-categories:
576 Amino acid metabolism, metabolism of terpenoids and polyketides, and biosynthesis of other
577 secondary metabolites.
578
579
580
581
582
583
584
585
586
587
588
589
590
591

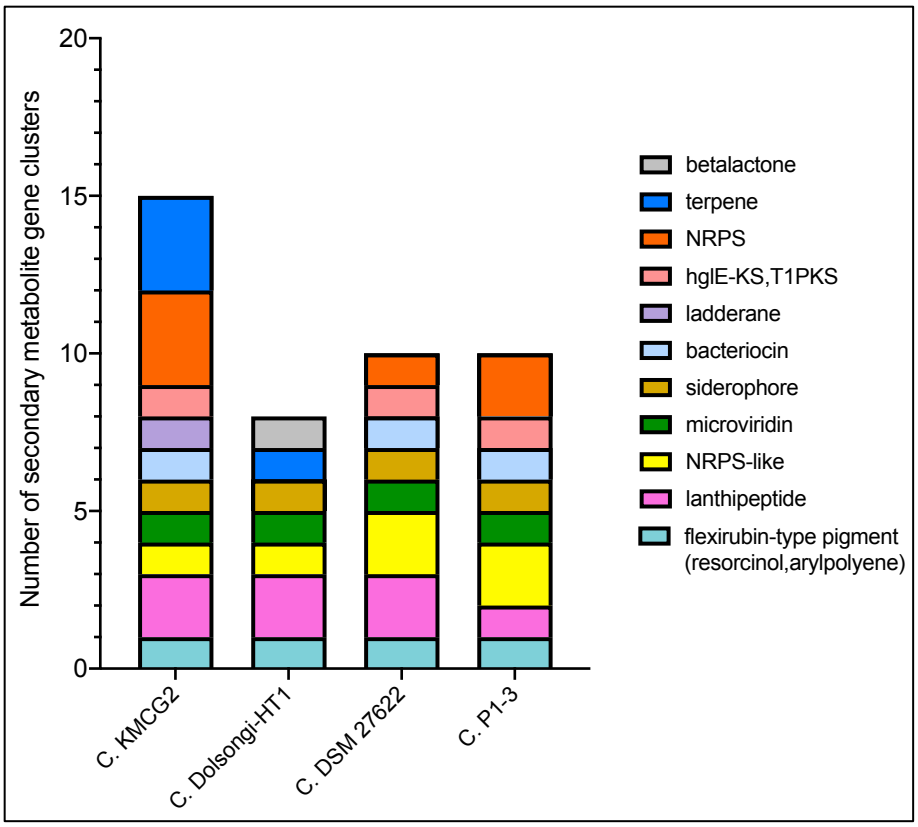
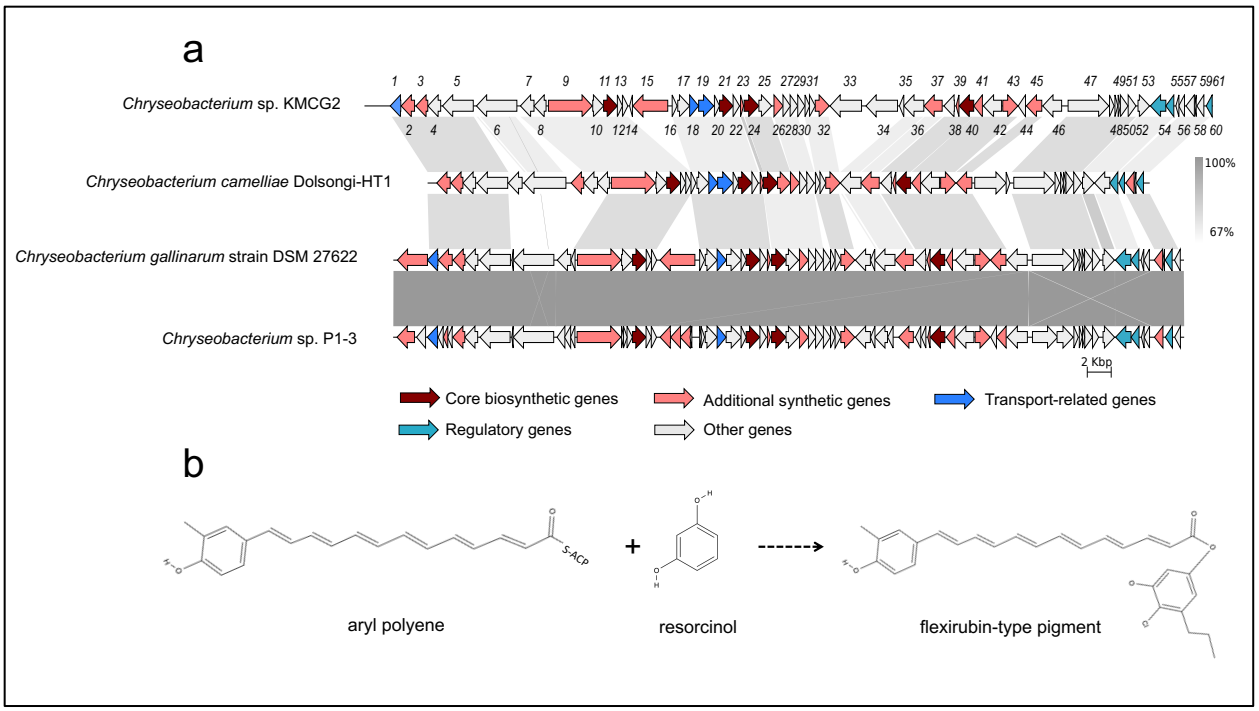


Fig. 2. Composition of secondary metabolite gene clusters from four *Chryseobacterium* spp. genomes.



605

606 **Fig. 3. Flexirubin-type pigment gene cluster from four *Chryseobacterium* spp. genomes.** (a).

607 Synteny analysis and features of flexirubin-type pigment gene cluster in *Chryseobacterium* spp.

608 genomes. (b). The proposed biosynthetic reaction of flexirubin-type pigment. The detailed

609 description of each gene can be found in Table S1.

610

611

612

613

614

615

616

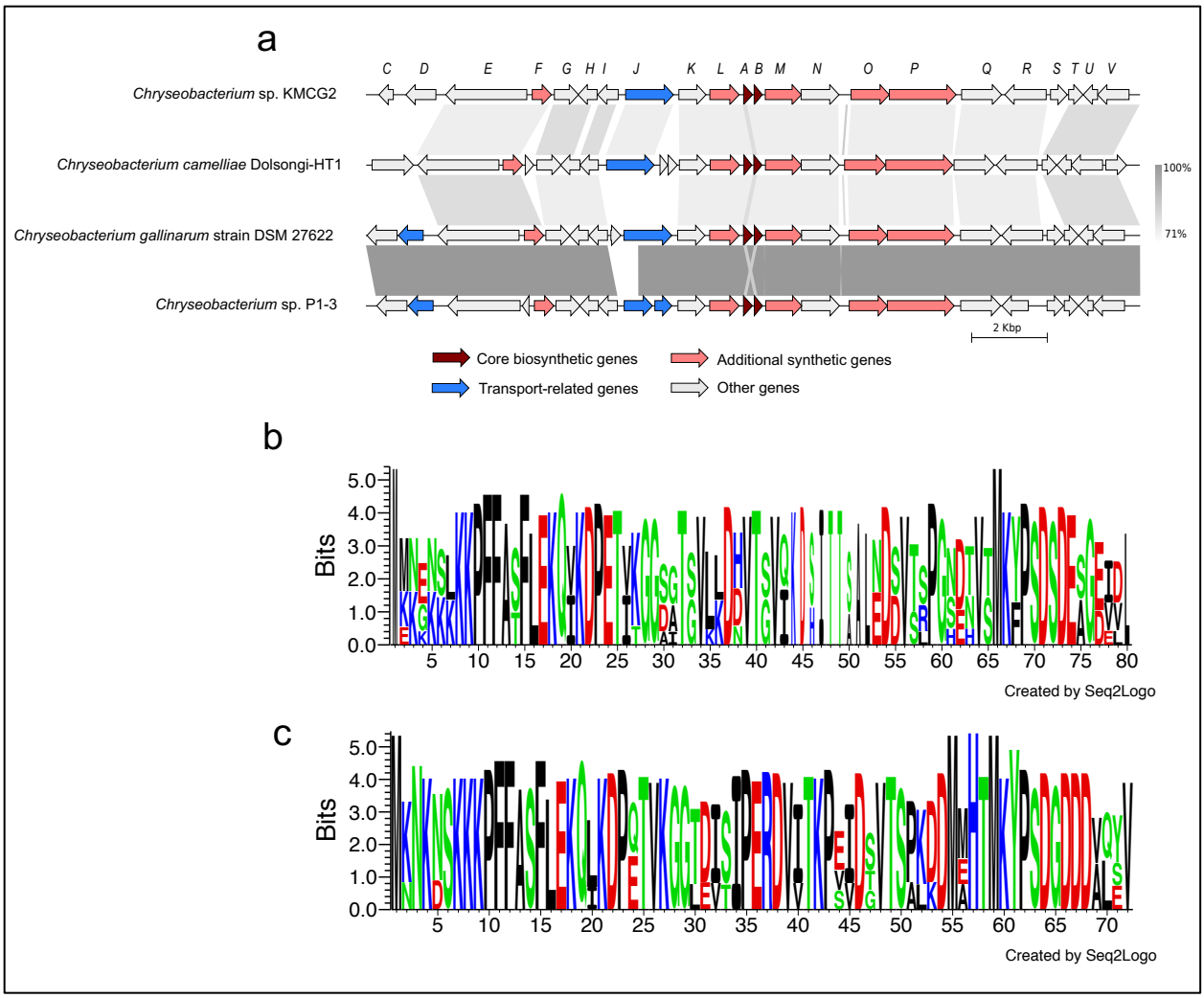


Fig. 4. Microviridin gene cluster from four *Chryseobacterium* spp. genomes. (a). Synteny analysis and features of microviridin gene cluster in *Chryseobacterium* spp. genomes. (b). Amino acid sequence comparison of mdnA from four *Chryseobacterium* spp. genomes. (c). Amino acid sequence comparison of mdnB from four *Chryseobacterium* spp. genomes. The detailed description of each gene can be found in Table S2.

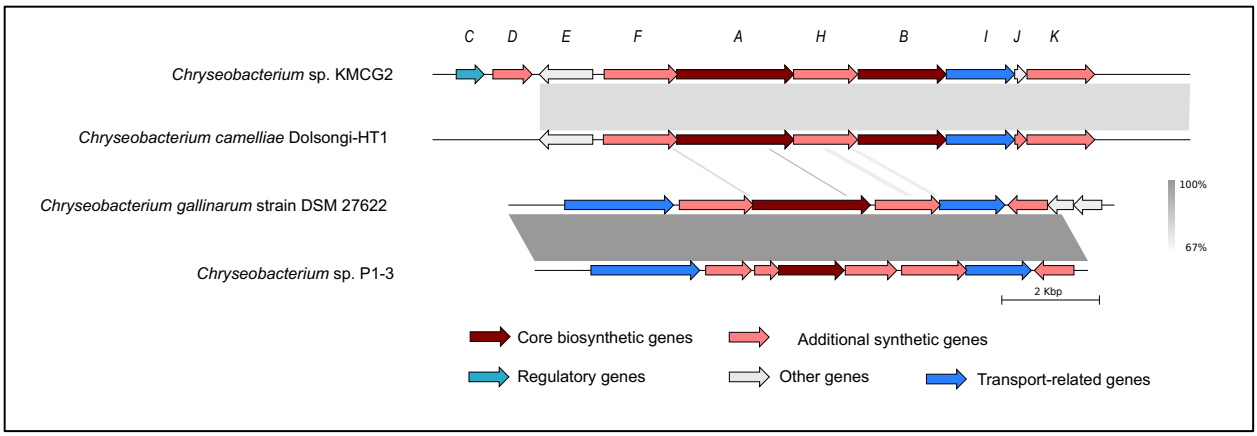
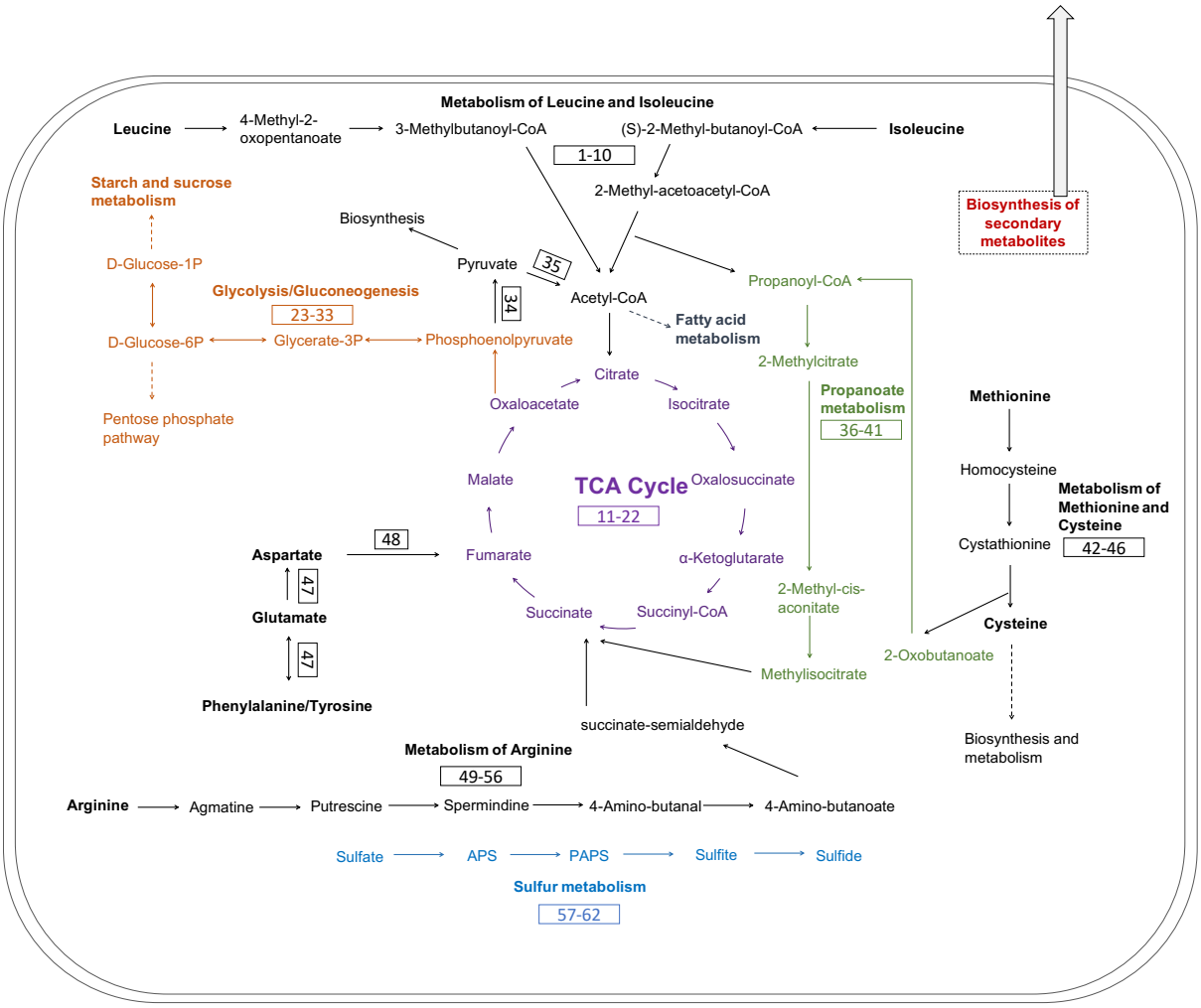


Fig. 5. Siderophore gene cluster from four *Chryseobacterium* spp. genomes. The detailed description of each gene can be found in Table S3.



643

644 **Fig. 6. Metabolic pathways reconstruction of keratin utilization in *Chryseobacterium* sp.**

645 **KMC2 genome.** It includes amino acid metabolism, TCA cycle, glycolysis/gluconeogenesis

646 propanoate metabolism, and sulfur metabolism. The number in the box represents the gene

647 related to the metabolic pathway. The detailed description of each gene can be found in Table S7.

648

649

650

651

652

Table 1. Features comparison of four *Chryseobacterium* spp. genomes.

Parameters	*Chryseobacterium* sp. KMC2	*Chryseobacterium camelliae* Dolsongi-HT1	*Chryseobacterium gallinarum* strain DSM 27622	*Chryseobacterium* sp. P1-3
Total length (bp)	5.276.159	4.376.354	4.633.632	4.628.764
Contigs	63	1	1	45
N50 (bp)	231.784	4.376.354	4.633.632	342.512
GC content (%)	36.33	41.80	37.30	37.02
Gene	4.773	3.999	4.182	4.087
CDS	4.692	3.881	4.033	3.119