



Evolutionary Relationships Between Dysregulated Genes in Oral Squamous Cell Carcinoma and Oral Microbiota

Yang Fang¹, Yi Yang² and Chengcheng Liu^{2*}

¹ Department of Laboratory Medicine, Third Affiliated Hospital of Zhengzhou University, Zhengzhou, China, ² State Key Laboratory of Oral Diseases, National Clinical Research Center for Oral Diseases, Department of Periodontics, West China School and Hospital of Stomatology, Sichuan University, Chengdu, China

OPEN ACCESS

Edited by:

Jin Xiao,
University of Rochester, United States

Reviewed by:

Xian Xia,
Hubei Normal University, China
Aruni Wilson,
Loma Linda University, United States

*Correspondence:

Chengcheng Liu
liuchengcheng519@163.com

Specialty section:

This article was submitted to
Microbiome in Health and Disease,
a section of the journal
Frontiers in Cellular and
Infection Microbiology

Received: 28 April 2022

Accepted: 20 June 2022

Published: 13 July 2022

Citation:

Fang Y, Yang Y and Liu C
(2022) Evolutionary Relationships
Between Dysregulated Genes
in Oral Squamous Cell Carcinoma
and Oral Microbiota.
Front. Cell. Infect. Microbiol. 12:931011.
doi: 10.3389/fcimb.2022.931011

Oral squamous cell carcinoma (OSCC) is one of the most prevalent cancers in the world. Changes in the composition and abundance of oral microbiota are associated with the development and metastasis of OSCC. To elucidate the exact roles of the oral microbiota in OSCC, it is essential to reveal the evolutionary relationships between the dysregulated genes in OSCC progression and the oral microbiota. Thus, we interrogated the microarray and high-throughput sequencing datasets to obtain the transcriptional landscape of OSCC. After identifying differentially expressed genes (DEGs) with three different methods, pathway and functional analyses were also performed. A total of 127 genes were identified as common DEGs, which were enriched in extracellular matrix organization and cytokine related pathways. Furthermore, we established a predictive pipeline for detecting the coevolutionary of dysregulated host genes and microbial proteomes based on the homology method, and this pipeline was employed to analyze the evolutionary relations between the seven most dysregulated genes (MMP13, MMP7, MMP1, CXCL13, CRISPO3, CYP3A4, and CRNN) and microbiota obtained from the eHOMD database. We found that cytochrome P450 3A4 (CYP3A4), a member of the cytochrome P450 family of oxidizing enzymes, was associated with 45 microbes from the eHOMD database and involved in the oral habitat of *Comamonas testosteroni* and *Arachnia rubra*. The peptidase M10 family of matrix metalloproteinases (MMP13, MMP7, and MMP1) was associated with *Lactocaseibacillus paracasei*, *Lactocaseibacillus rhamnosus*, *Streptococcus salivarius*, *Tannerella* sp._HMT_286, and *Streptococcus infantis* in the oral cavity. Overall, this study revealed the dysregulated genes in OSCC and explored their evolutionary relationship with oral microbiota, which provides new insight for exploring the microbiota–host interactions in diseases.

Keywords: OSCC, oral microbiota, evolutionary relationships, matrix metalloproteinases, DEGs1

Abbreviations: DEGs, differentially expressed genes; MMPs, peptidase M10 family of matrix metalloproteinases; OSCC, oral squamous cell carcinoma; GEO, Gene Expression Omnibus; GO, The Gene Ontology; MF, Molecular Function; CC, Cellular Component; BP, Biological Process; FDR, adjusted P-value; FC: fold change; KEGG, Kyoto Encyclopedia of Genes and Genomes.

INTRODUCTION

Head and neck squamous cell carcinoma (HNSCC) is the seventh most common malignancy in the world, accounting for more than 90% of head and neck malignancies (Mody et al., 2021). HNSCC originates from the squamous epithelium of the upper respiratory tract and digestive tract of the oral cavity, pharynx, and larynx, among which pharyngeal squamous cell carcinoma, laryngeal squamous cell carcinoma, and OSCC are the most common. OSCC often has a great impact on patients' chewing, swallowing, language, breathing, and other functions and even threatens their lives (Johnson et al., 2020). In recent years, the incidence of OSCC has been on the rise, becoming a world public health problem with high morbidity and mortality (Ferlay et al., 2019). Studies have shown that, in addition to major risk factors such as tobacco and alcohol abuse, exposure to environmental pollutants and viruses, specific oral bacteria, or oral microbial communities may play an important role in the occurrence and progression of OSCC (Fitzsimonds et al., 2020; Irfan et al., 2020). The human microbiome coevolved and coexisted, and OSCC that also grows in the oral cavity may themselves be the hosts of oral microbiota. The oral cavity harbors over 700 microbial species and both pathogenic and commensal strains are involved in the development of OSCC. Evidence has indicated a correlation of some specific species with OSCC, including *Porphyromonas gingivalis*, *Fusobacterium nucleatum*, *Treponema denticola*, *Streptococcus gordonii*, and human papilloma virus 16 (Fitzsimonds et al., 2020).

The current strategies to investigate the role of the oral microbiota in OSCC have predominantly focused on detecting oral microbial communities present or populational shifts in OSCC samples and studying the effect and mechanism of specific oral microbial challenges on biological processes (BPs) related to OSCC occurrences, such as cell proliferation, cell apoptosis, and the epithelial to mesenchymal transition. Perspective studies should focus on exploring the oral microbiota potentially related to OSCC. However, oral microbes are abundant and approximately 30% of them cannot be cultured. Thus, determining oral microbial–host interactions between species experimentally is a challenging task (Fritz et al., 2013). Computational approaches are an ideal approach to aid in screening for microbial–host interactions, with time-saving and economic advantages (Dix et al., 2016; Fitzsimonds et al., 2020). From an evolutionary perspective, if there is significant similarity between two protein sequences, they may originate from a common ancestor and have the same or similar functions (Pearson, 2013). Therefore, the most common method to explore protein function is pair-wise protein sequence comparison to “transfer” or prediction of function based on sequence similarity between proteins of known and unknown function (Rost, 1999; Bork, 2000; Devos and Valencia, 2000). And BLAST is a classic pairwise approach that can search protein sequence similarities all against all (Camacho et al., 2009). Based on this principle, several approaches have been established to determine pathogen–host protein–protein interactions (PHIs), including protein homology prediction, structural domain-based methods, and machine learning-based methods (English

and Albersheim, 1969; Wojcik and Schächter, 2001; Pagel et al., 2004; Qi et al., 2010; Dyer et al., 2011).

Increasing evidence has shown that shared evolutionary history matters to both microbiota and hosts (Davenport et al., 2017). Microorganisms include bacteria, viruses, fungi, and some small protists. These organisms typically have smaller proteomes, based on which they can be analyzed for community heterogeneity, activity, and function. If the entire proteome of a microorganism is regarded as a protein that is evolutionarily conserved, the entire microbiome is a living organism composed of a large number of such proteins. Based on the principle of protein–protein interaction (PPI), the interaction relationship between the entire proteome of microorganisms and the host can be predicted. Oral microbiota colonize the oral mucosa, and they have a coevolutionary relationship; therefore, based on the principle of coevolutionary association inferring functional interactions, this study established a predictive pipeline for the evolutionarily interconnected evolution of dysregulated host genes and microbial proteomes (Devos and Valencia, 2000). We identified dysregulated genes in OSCC tissues by analyzing gene expression datasets from the GEO database and explored their evolutionary relations with oral microbiota with this pipeline. The results are expected to provide new insights into the interactions between oral microbiota and OSCC.

MATERIALS AND METHODS

Data Collection

The Gene Expression Omnibus (GEO) database contains a large number of gene expression profiling (high-throughput sequencing and microarray datasets) and RNA methylation profiles that are submitted by different research laboratories in the world (Edgar et al., 2002). We retrieved the related gene expression datasets by OSCC and oral keywords. The criteria for the retrieved datasets must contain different sequencing platforms. Finally, the microarray datasets (GSE138206) and high-throughput sequencing datasets (GSE140707) were downloaded from the GEO database. The GSE138206 dataset contains six OSCC tissues (Ca), tissues adjacent to cancer (P), and contralateral normal tissues (N), and the GSE140707 contains three tumorous and adjacent tissues from OSCC sufferers.

DEGs Identification

All analyses of differentially expressed genes (DEGs) were performed by R language. The GEO query package (Sean and Meltzer, 2007) was used to obtain collected datasets from the GEO database. For microarray datasets, the statistically significant DEGs were acquired by utilizing the limma package (Ritchie et al., 2015) with adjusted P-value (FDR) < 0.05 and $|\log_2(\text{fold change (FC)})| > 1$. The Deseq2 (Love et al., 2014) and EdgeR (Robinson et al., 2010) packages were utilized to analyze RNA sequencing data and filter significant DEGs between OSCC and adjacent normal tissues.

Enrichment Analyses

The Gene Ontology (GO) describes our knowledge of the biological function in three aspects: molecular function (MF), cellular component (CC), and BP (The Gene Ontology, Consortium 2019). The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway is a database utilized for genomic and biological pathway and other omics studies (Kanehisa and Goto, 2000). We used the ClusterProfile package (Yu et al., 2012) to identify the potential functions of the significant DEGs with the GO and KEGG databases, and the results were displayed by the ggplot2 package (Wickham, 2011) with a cutoff of $p < 0.05$.

Evolutionary Relation With the Oral Microbiome

The microbe data (1,903 microbiome genomes with 4,665,857 proteins) were downloaded from the expanded Human Oral Microbiome Database (eHOMD) (Escapa et al., 2018). The eHOMD provides comprehensive curated information on bacteria in the human mouth and aerodigestive tract, including the pharynx, nasal passages, sinuses, and esophagus. The DEG homologous proteins were searched by BLAST+ software (Camacho et al., 2009) with an e-value < 0.001 . We used the MEGA11 software to identify the evolutionary relationship between DEGs and oral microbiomes by MEGA11 software (Tamura et al., 2021). Multiple sequence alignment was performed by MUSCLE (Edgar, 2004) (Gap Open:-2.9, Hydrophobicity Multiplier:1.2, Max Iterations:16, Min Diag Length:24), using the maximum likelihood method (JTT model and NNI ML heuristic method) to reconstruct the phylogenetic tree. The iTOL is used to illustrate the phylogenetic tree (Letunic and Bork, 2006).

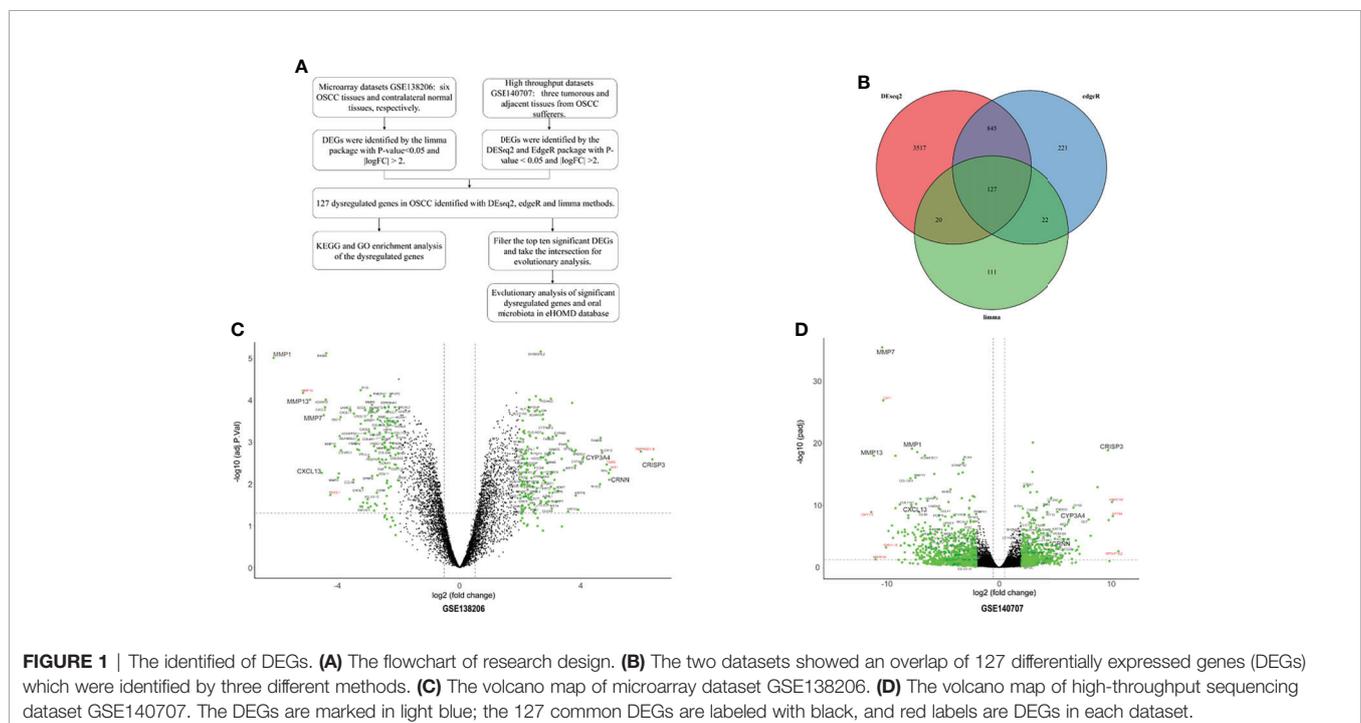
RESULTS

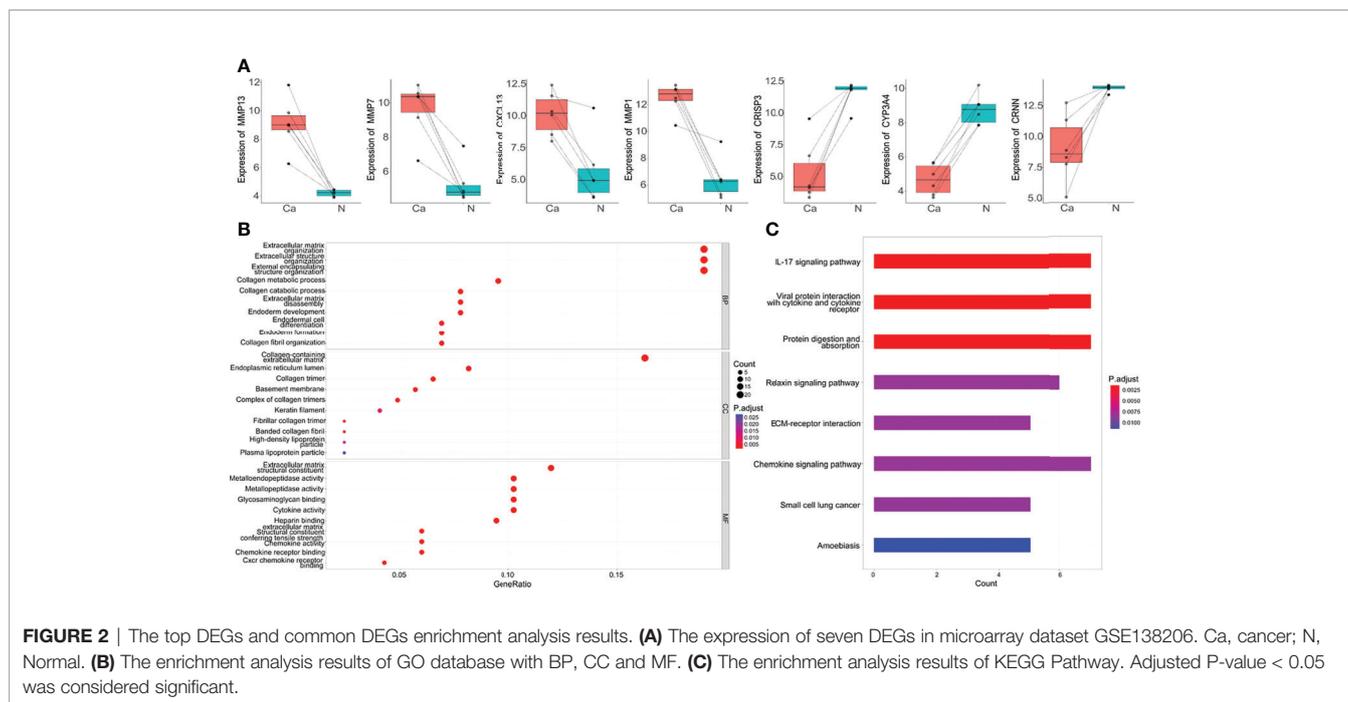
Transcriptional Landscapes of OSCC

This study was performed according to the workflow (Figure 1A). A total of 280 DEGs in microarray dataset GSE138206 were identified by the limma package with 162 upregulated genes and 118 downregulated genes (Figure 1C, Table S1). Analysis of the high-throughput sequencing datasets GSE140707 by Deseq2 packages obtained 1,699 DEGs of OSCC, and, of these, 652 genes and 1,047 genes were upregulated and downregulated, respectively (Figure 1D, Table S2). In addition, 1,215 DEGs were identified by EdgeR with 666 upregulated genes and 549 downregulated genes (Table S3). Taking the intersection of three different package analyses of two OSCC datasets and plotting the Venn diagram, 127 common DEGs were obtained (Figure 1B, Table S4). Additionally, there were five significant DEGs (MMP10, MUCL1, TGM3, WIF1, and TMPRSS11B) that were identified only in GSE138206 (Figure 1C). In addition to the intersection, seven high fold change expression genes (CST1, IGHV1-3, IGHV1-18, MAGEA6, HMGCS2, KRT84, and KRTAP13-2) were identified in GSE140707 (Figure 1D).

Dysregulation of Genes Related to Extracellular Matrixes and Cytokines in OSCC

Seven genes belonged to the intersection of DEGs identified by three different methods were dysregulated genes in OSCC. Among them, MMP13, MMP7, MMP1, and CXCL13 genes were upregulated in OSCC tissues, and CRISP3, CYP3A4, and CRNN genes were downregulated (Figure 2A). To gain insight into the pathways and function of common DEGs of OSCC, the





GO enrichment analysis and KEGG enrichment analysis were performed. It was observed that DEGs were enriched in categories associated with extracellular matrix (ECM) organization, collagen metabolic process, metalloproteinase activity, glycosaminoglycan binding, and cytokine activity (Figure 2B). KEGG analysis showed that the 127 DEGs were significantly enriched in 10 pathways, such as cytokine-cytokine receptor interaction, IL-17 signaling pathway, viral protein interaction with cytokine and cytokine receptor, and protein digestion and absorption. Notably, COL4A1, COL4A2, COL4A6, FN1, and LAMC2 genes were also enriched in the ECM receptor interaction pathway (Figure 2C).

Evolutionary Relationship Between the Dysregulated Genes and Oral Microbiota

To explore the evolutionary relationships of the seven DEGs and oral microbiota. We used the BLAST method to align 1,903

microbiome genomes (total 4,665,857 proteins) in human microbiome which retrieved from the eHOMD database. We set the BLAST cutoff with e-value < 10⁻³ against the microbiome genome search for the homologous proteins (Table 1). We searched 350 homologous proteins within 45 species of microorganisms with the CYP3A4 gene by BLAST software. Among these microorganisms, *Comamonas testosteroni* KF-1, *Comamonas testosteroni* CNB-2, *Comamonas testosteroni* S44, and *Arachnia rubra* DSMZ 10012 inhabit in the oral cavity (Table S5). Oral colonizers with an evolutionary relationship to MMP1 include *Streptococcus infantis* ATCC 700779, *Streptococcus infantis* SPAR10, and *Tannerella* sp._HMT_286 W11667 (Table S6). Interestingly, *Tannerella* sp._HMT_286 W11667 and MMP5 as well as MMP13 also have a coevolutionary relationship (Tables S7, S8). In addition, MMP13 also has a coevolutionary relationship with *Lactacisbacillus paracasei*, *Lactacisbacillus rhamnosus*, and

TABLE 1 | The homologous proteins searched of seven DEGs.

Gene name	Gene annotation	Species	Strain	Oral	Uassigned	Protein	Species name
CRISP3	cysteine-rich secretory protein (CRISP) family	1	1	0	1	1	NA
CYP3A4	cytochrome P450 superfamily of enzymes	45	119	2	43	350	<i>Comamonas testosteroni</i> , <i>rubra</i>
MMP1	peptidase M10 family of matrix metalloproteinases	7	39	4	3	40	<i>Lactacisbacillus rhamnosus</i> , <i>Lactacisbacillus paracasei</i> , <i>Streptococcus salivarius</i> , <i>Tannerella</i> sp._HMT_286
MMP7	peptidase M10 family of matrix metalloproteinases	9	16	1	8	16	<i>Tannerella</i> sp._HMT_286
MMP13	peptidase M10 family of matrix metalloproteinases	5	11	2	3	13	<i>Streptococcus infantis</i> , <i>Tannerella</i> sp._HMT_286
CXCL13	C-X-C Motif Chemokine Ligand 13	0	0	0	0	0	NA
CRNN	The "fused gene" family of proteins	0	0	0	0	0	eNA

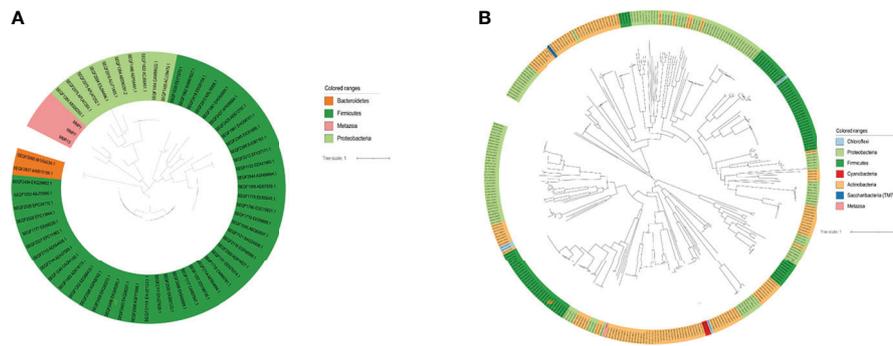


FIGURE 3 | The evolutionary relation analysis of Top DEGs. **(A)** The evolutionary relation between MMP13, MMP7, and MMP1 with the microbiome. **(B)** The evolutionary relation between CYP3A4 with the microbiome.

Streptococcus salivarius, which live in the oral cavity (Table S8). Figure 3A shows the evolutionary relations among the MMP13, MMP7, and MMP1 genes of oral genomes. As can be seen from the results of the analysis of the evolutionary relationship, three upregulated genes have homologs with the phyla of *Bacteroidetes*, *Firmicutes*, and *Proteobacteria*. The most significant aspect of this relationship is *Proteobacteria*, which contains 8 oral species and 10 proteins. The evolutionary relations of CYP3A4 genes are shown in Figure 3B.

DISCUSSION

OSCC is the most common malignancy, accounting for 80%–90% of oral malignancies. Oral microbiota is a major risk factor for OSCC. Associated interactions between oral microorganisms and host can promote the progression of OSCC (Bai et al., 2022). In this study, we identified 652 upregulated and 1,047 downregulated genes in the OSCC tissues based on the GSE140707 dataset, as well as 162 upregulated and 118 downregulated genes in the GSE138216 dataset. The unique significance of 127 DEGs by three methods was based on the analyzed metadata. DEGs were identified in both datasets associated with the IL-17 signaling pathway, viral protein interaction with cytokines and cytokine receptors, and protein digestion and absorption. Seven significant dysregulated genes in OSCC tissues were further identified, including four upregulated genes MMP13, MMP7, MMP1, and CXCL13 and three downregulated genes CRSP3, CYP3A4, and CRNN. In the present study, we established a predictive pipeline for exploring the evolutionary relationship between the oral microbiota in the eHOMD database and the dysregulated genes in OSCC based on the principle of coevolution.

Matrix metalloproteinases (MMPs) are involved in normal physiological processes of decomposing ECM, such as tissue remodeling, embryonic development, and reproduction, as well as in disease processes of arthritis and metastasis (Snoek-van Beurden and Von den Hoff, 2005). The MMP13, MMP7, and MMP1 genes encode members of the peptidase M10 family of

MMPs. Consistent with our findings, previous studies have reported the dysregulated expression of them in OSCC progression. For instance, overexpression of both transcriptional and translational levels of MMP13 was found in OSCC tissues (Johansson et al., 1997; Culhaci et al., 2004; Luukka et al., 2006). Highly expressed MMP13 protein also showed a significant correlation with tumor staging and lymph node metastasis (Vincent-Chong et al., 2014). The expression of MMP7 and MMP1 was also upregulated in OSCC tissues. The expression of MMP7 and MMP1 were also upregulated in OSCC tissues. The MMP1 gene was activated in aggressive OSCC (Impola et al., 2004; Jordan et al., 2004; Chuang et al., 2008; Mäkinen et al., 2014).

Moreover, the MMP1 gene might be used as a potential target to improve diagnosis and as an oral cancer marker for OSCC (Hashimoto et al., 2004; Yen et al., 2009; Yang et al., 2020). Functional enrichment analysis revealed that MMP13, MMP7, and MMP1 were related to ECM organization, ECM disassembly, extracellular structure organization, and endoderm development. Tumors can utilize ECM remodeling to create a microenvironment that promotes tumorigenesis and metastasis. Therefore, we speculate that the overexpression of MMP13, MMP7, and MMP1 might involve invasiveness and metastasis of OSCC by modulating ECM remodeling.

The homology search method identified an evolutionary relationship with the oral microbiota for the MMP13, MMP7, and MMP1 genes. We chose MMP family genes to explore their evolutionary relationship with microbial species because MMP family differential expression was significant when differentially expressed genes were analyzed. The *Lactocaseibacillus paracasei*, *Lactobacillus rhamnosus*, *Streptococcus salivarius*, *Tannerella* sp. HMT 286, and *Streptococcus infantis* five species were hit homologous. In the previous study, Pushalkar et al. assessed the microbial diversity in OSCC tissues and non-tumor tissues. The results showed that the microbial load of *Lactocaseibacillus paracasei* has a significant variation (Pushalkar et al., 2012). It has been reported that the administration of *Lactobacillus rhamnosus* was able to increase the effect of anticancer molecules tested on human OSCC (Cheng et al., 2017). These results

demonstrated the potential of *Lactobacillus rhamnosus* as a beneficial effect adjuvant treatment for OSCC. In OSCC patients undergoing tumor resection, the percentage of saliva-reactive cytotoxic T cells was positively correlated with recurrence-free survival (Wang et al., 2018). Another study (Pavlova et al., 2013) showed that *Streptococcus salivarius* is involved in alcohol metabolism to acetaldehyde, which has a carcinogenic potential (Vogelmann and Amieva, 2007; Marttila et al., 2013). Consistent with our findings, evidence has also shown a significant difference of *Streptococcus infantis* between OSCC patients and healthy individuals (Hsiao et al., 2018). These all support the reliability of our predictive pipeline for exploring the interaction between dysregulated genes in OSCC and oral microbiota.

The CYP3A4 gene encodes a member of the cytochrome P450 superfamily of enzymes and is involved in the metabolism of sterols, steroid hormones, retinoids, and fatty acids (Chen et al., 2000; Marill et al., 2000; Badawi et al., 2001). Cytoscape software by the Matthews correlation coefficient (MCC) algorithm was used to predict CYP3A4 at the core position in the network and highlight the first 10 types of OSCC DEGs (Li et al., 2022). It is worth noting that there found 350 homologous proteins including 45 microbes were found by evolutionary analysis in this study. Among them, *Comamonas testosteroni* and *Arachnia rubra* can inhabit in the oral cavity. Up to 348 microorganisms were unassigned information on their location in the human body. Extracts of *Arachnia rubra* are associated with human OSCC production and modulation of tumor-specificity values (Suzuki et al., 2014). However, we noticed that the two previously reported bacteria associated with OSCC, *Porphyromonas gingivalis* and *Clostridium perfringens*, were not found when using this method to explore the relationship between dysregulated genes and oral microbiota. We speculate that the possible reason is that, although these bacteria play an important role in OSCC, there is no homologous evolutionary relationship with the genes that we screened. CXCL13, CRISPO3, and CRNN have newly identified dysregulated genes by this study, and their effects on the OSCC need to be further investigated for experimental validation.

Microorganisms, including bacteria, viruses, and archaea, inhabit a wide range of hosts in different ecological niches and ecosystems (Braga et al., 2016). Deciphering microbial–host interactions can provide new therapeutic strategies for maintaining health or improving disease states. However, determining microbial–host interactions between species experimentally is a challenging task due to many other limitations related to the size, scope, feasibility of studies, and sample availability of microbial populations (Fritz et al., 2013). Computational approaches can overcome some of these limitations and thus enhance our understanding of microbial–host interactions (Dix et al., 2016). Molecular ecological networks are used to study the interactions between molecules (from different species or even kingdoms) in a larger ecosystem (Yang et al., 2017; Meyer et al., 2020). From a mechanistic perspective, the most widely studied types of interactions among species interactions include microbial networks, PPIs, and RNA-mediated interactions. Therefore, many

computational methods developed to study microbe–host interactions have focused on the three types of interactions mentioned above. However, all of these inference methods have the feature of studying microbe–host interactions based on the characteristics of the sequence structure. In this study, we propose the use of coevolutionary principles to infer microbial–host interactions based on sequence structure. The use of the coevolutionary principle better reflects the conserved protein structure of the species than the direct use of sequence structure. Therefore, we established a predictive pipeline to study the interaction between DEGs and oral microbiota in OSCC based on sequence-structure conservativeness and coevolutionary principles. Of course, the method has its shortcomings, that is, the method is based on the inference that species have the same protein conserved modules during the evolutionary process, and if the studied DEGs and microbial proteins do not have the same evolutionary conserved modules, they cannot be studied by this method. However, this method provides a novelty way and new ideas for exploring the relationship between host genes and host symbiotic microorganism.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

YF contributed to the study conception and design, data acquisition, analysis, and interpretation, and in the drafting the of the manuscript. YF and YY contributed to data acquisition, analysis, and interpretation. CL contributed to the study conception and design and the drafting and critical revision of the manuscript. All the authors approved the final version of the manuscript and agreed to be accountable for all aspects of the work.

FUNDING

This work is supported by PhD research startup foundation of the Third Affiliated Hospital of Zhengzhou University (2021080).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcimb.2022.931011/full#supplementary-material>

Supplementary Table 1 | The DEGs in GSE138206 dataset were identified by limma packages.

Supplementary Table 2 | The DEGs in GSE140707 dataset were identified by the Deseq2 package.

Supplementary Table 3 | The DEGs in GSE140707 dataset were identified by the EdgeR package.

Supplementary Table 4 | The 127 common DEGs.

Supplementary Table 5 | The homolog searched results of CYP3A4 genes.

Supplementary Table 6 | The homolog searched results of MMP13 genes.

Supplementary Table 7 | The homolog searched results of MMP7 genes.

Supplementary Table 8 | The homolog searched results of MMP1 genes.

REFERENCES

- Badawi, A. F., Cavalieri, E. L., and Rogan, E. G. (2001). "Role of Human Cytochrome P450 1a1, 1A2, 1B1, and 3A4 in the 2-, 4-, and 16 Alpha-Hydroxylation of 17 Beta-Estradiol." *Metab. Clin. Exp.* 50 (9), 1001–1003. doi: 10.1053/meta.2001.25592
- Bai, H., Yang, J., Meng, S., and Liu, X. X. C. (2022). "Oral Microbiota-Driven Cell Migration in Carcinogenesis and Metastasis." *Front. Cell Infect. Microbiol.* 12. doi: 10.3389/fcimb.2022.864479
- Bork, P. (2000). "Powers and Pitfalls in Sequence Analysis: The 70% Hurdle." *Genome Res.* 10 (4), 398–400. doi: 10.1101/gr.10.4.398
- Braga, R. M., Dourado, M. N., and Araujo, W. L. (2016). "Microbial Interactions: Ecology in a Molecular Perspective." *Braz. J. Microbiol.* 47 (Suppl 1), 86–98. doi: 10.1016/j.bjm.2016.10.005
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). "BLAST+: Architecture and Applications." *BMC Bioinf.* 10, 421. doi: 10.1186/1471-2105-10-421
- Cheng, Z., Xu, H., Wang, X., and Liu, Z. (2017). "Lactobacillus Raises *In Vitro* Anticancer Effect of Geniposide in HSC-3 Human Oral Squamous Cell Carcinoma Cells." *Exp. Ther. Med.* 14 (5), 4586–4594. doi: 10.3892/etm.2017.5105
- Chen, H., Howald, W. N., and Juchau, M. R. (2000). "Biosynthesis of All-Trans-Retinoic Acid From All-Trans-Retinol: Catalysis of All-Trans-Retinol Oxidation by Human P-450 Cytochromes." *Drug Metab. Disposition* 28 (3), 315–322.
- Chuang, H. C., Su, C. Y., Huang, H. Y., Huang, C. C., Chien, C. Y., Du, Y. Y., et al. (2008). "Active Matrix Metalloproteinase-7 is Associated With Invasion in Buccal Squamous Cell Carcinoma." *Modern Pathol.* 21 (12), 1444–1450. doi: 10.1038/modpathol.2008.99
- Culhaci, N., Metin, K., Copcu, E., and Dikicioglu, E. (2004). "Elevated Expression of MMP-13 and TIMP-1 in Head and Neck Squamous Cell Carcinomas may Reflect Increased Tumor Invasiveness." *BMC Cancer* 4, 42. doi: 10.1186/1471-2407-4-42
- Davenport, E. R., Sanders, J. G., Song, S. J., Amato, K. R., Clark, A. G., and Knight, R. (2017). "The Human Microbiome in Evolution." *BMC Biol.* 15 (1), 127. doi: 10.1186/s12915-017-0454-7
- Devos, D., and Valencia, A. (2000). "Practical Limits of Function Prediction." *Proteins Struct. Funct. Bioinf.* 41 (1), 98–107. doi: 10.1002/1097-0134(20001001)41:1<98::AID-PROT120>3.0.CO;2-S
- Dix, A., Vlaic, S., Guthke, R., and Linde, J. (2016). "Use of Systems Biology to Decipher Host-Pathogen Interaction Networks and Predict Biomarkers." *Clin. Microbiol. Infect.* 22 (7), 600–606. doi: 10.1016/j.cmi.2016.04.014
- Dyer, M. D., Murali, T. M., and Sobral, B. W. (2011). "Supervised Learning and Prediction of Physical Interactions Between Human and HIV Proteins." *Infect. Genet. Evol.* 11 (5), 917–923. doi: 10.1016/j.meegid.2011.02.022
- Edgar, R. C. (2004). "MUSCLE: Multiple Sequence Alignment With High Accuracy and High Throughput." *Nucleic Acids Res.* 32 (5), 1792–1797. doi: 10.1093/nar/gkh340
- Edgar, R., Domrachev, M., and Lash, A. E. (2002). "Gene Expression Omnibus: NCBI Gene Expression and Hybridization Array Data Repository." *Nucleic Acids Res.* 30 (1), 207–210. doi: 10.1093/nar/30.1.207
- English, P. D., and Albersheim, P. (1969). Host-Pathogen Interactions: I. A Correlation Between α -Galactosidase Production and Virulence I. *Plant Physiol.* 44 (2), 217–224. doi: 10.1104/pp.44.2.217
- Escapa, I. F., Chen, T., Huang, Y., Gajare, P., Dewhirst, F. E., and Lemon, K. P. (2018). "New Insights Into Human Nostril Microbiome From the Expanded Human Oral Microbiome Database (eHOMD): A Resource for the Microbiome of the Human Aerodigestive Tract." *Msystems* 3 (6), e00187–18. doi: 10.1128/mSystems.00187-18
- Ferlay, J., Colombet, M., Soerjomataram, I., Mathers, C., Parkin, D. M., Pineros, M., et al. (2019). "Estimating the Global Cancer Incidence and Mortality in 2018: GLOBOCAN Sources and Methods." *Int. J. Cancer* 144 (8), 1941–1953. doi: 10.1002/ijc.31937
- Fitzsimonds, Z. R., Rodriguez-Hernandez, C. J., Bagaitkar, J., and Lamont, R. J. (2020). "From Beyond the Pale to the Pale Riders: The Emerging Association of Bacteria With Oral Cancer." *J. Dental Res.* 99 (6), 604–612. doi: 10.1177/0022034520907341
- Fritz, J. V., Desai, M. S., Shah, P., Schneider, J. G., and Wilmes, P. (2013). "From Meta-Omics to Causality: Experimental Models for Human Microbiome Research." *Microbiome* 1 (1), 14. doi: 10.1186/2049-2618-1-14
- Hashimoto, T., Uchida, K., Okayama, N., Imate, Y., Suehiro, Y., Hamanaka, Y., et al. (2004). "Association of Matrix Metalloproteinase (MMP)-1 Promoter Polymorphism With Head and Neck Squamous Cell Carcinoma." *Cancer Lett.* 211 (1), 19–24. doi: 10.1016/j.canlet.2004.01.032
- Hsiao, J. R., Chang, C. C., Lee, W. T., Huang, C. C., Ou, C. Y., Tsai, S. T., et al. (2018). "The Interplay Between Oral Microbiome, Lifestyle Factors and Genetic Polymorphisms in the Risk of Oral Squamous Cell Carcinoma." *Carcinogenesis* 39 (6), 778–787. doi: 10.1093/carcin/bgy053
- Impola, U., Uitto, V. J., Hietanen, J., Hakkinen, L., Zhang, L., Larjava, H., et al. (2004). "Differential Expression of Matrilysin-1 (MMP-7), 92 kD Gelatinase (MMP-9), and Metalloelastase (MMP-12) in Oral Verrucous and Squamous Cell Cancer." *J. Pathol.* 202 (1), 14–22. doi: 10.1002/path.1479
- Irfan, M., Delgado, R. Z. R., and Frias-Lopez, J. (2020). "The Oral Microbiome and Cancer." *Front. Immunol.* 11. doi: 10.3389/fimmu.2020.591088
- Johansson, N., Airoola, K., Grénman, R., Kariniemi, A. L., Saarialho-Kere, U., and Kähäri, V. M. (1997). "Expression of Collagenase-3 (Matrix Metalloproteinase-13) in Squamous Cell Carcinomas of the Head and Neck." *Am. J. Pathol.* 151 (2), 499–508.
- Johnson, D. E., Burtneis, B., René Leemans, C., Lui, V. W. Y., Bauman, J. E., and Grandis, J. R. (2020). Head and Neck Squamous Cell Carcinoma. *Nat. Rev. Dis. Primers* 6 (1), 92. doi: 10.1038/s41572-020-00224-3
- Jordan, R. C. K., Macabeo-Ong, M., Shiboski, C. H., Dekker, N., Ginzinger, D. G., Wong, D. T. W., et al. (2004). "Overexpression of Matrix Metalloproteinase-1 and-9 mRNA is Associated With Progression of Oral Dysplasia to Cancer." *Clin. Cancer Res.* 10 (19), 6460–6465. doi: 10.1158/1078-0432.Ccr-04-0656
- Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28 (1), 27–30. doi: 10.1093/nar/28.1.27
- Letunic, I., and Bork, P. (2006). Interactive Tree Of Life (iTOL): An Online Tool for Phylogenetic Tree Display and Annotation. *Bioinformatics* 23 (1), 127–128. doi: 10.1093/bioinformatics/btl529
- Li, C. S., Shi, Y. Y., Zuo, L. H., Xin, M. Z., Guo, X. M., Sun, J. L., et al. (2022). "Identification of Biomarkers Associated With Cancerous Change in Oral Leukoplakia Based on Integrated Transcriptome Analysis." *J. Oncol.* 2022, 4599305. doi: 10.1155/2022/4599305
- Love, M. I., Huber, W., and Anders, S. (2014). "Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data With Deseq2." *Genome Biol.* 15 (12), 550. doi: 10.1186/s13059-014-0550-8
- Luukka, M., Vihinen, P., Kronqvist, P., Vahlberg, T., Pyrhönen, S., Kähäri, V. M., et al. (2006). "Association Between High Collagenase-3 Expression Levels and Poor Prognosis in Patients With Head and Neck Cancer." *Head Neck* 28 (3), 225–234. doi: 10.1002/hed.20322
- Makinen, L. K., Hayry, V., Hagstrom, J., Sorsa, T., Passador-Santos, F., Keski-Santti, H., et al. (2014). "Matrix Metalloproteinase-7 and Matrix Metalloproteinase-25 in Oral Tongue Squamous Cell Carcinoma." *Head Neck J. Sci. Specialties Head Neck* 36 (12), 1783–1788. doi: 10.1002/hed.23539
- Marill, J., Cresteil, T., Lanotte, M., and Chabot, G. G. (2000). "Identification of Human Cytochrome P450s Involved in the Formation of All-Trans-Retinoic

- Acid Principal Metabolites." *Mol. Pharmacol.* 58 (6), 1341–1348. doi: 10.1124/mol.58.6.1341
- Marttila, E., Bowyer, P., Sanglard, D., Uittamo, J., Kaihovaara, P., Salaspuro, M., et al. (2013). "Fermentative 2-Carbon Metabolism Produces Carcinogenic Levels of Acetaldehyde in *Candida Albicans*." *Mol. Oral. Microbiol.* 28 (4), 281–291. doi: 10.1111/omi.12024
- Meyer, J. M., Leempoel, K., Losapio, G., and Hadly, E. A. (2020). "Molecular Ecological Network Analyses: An Effective Conservation Tool for the Assessment of Biodiversity, Trophic Interactions, and Community Structure." *Front. Ecol. Evol.* 8. doi: 10.3389/fevo.2020.588430
- Mody, M. D., Rocco, J. W., Yom, S. S., Haddad, R. I., and Saba, N. F. (2021). "Head and Neck Cancer." *Lancet* 398 (10318), 2289–2299. doi: 10.1016/s0140-6736(21)01550-6
- Pagel, P., Wong, P., and Frishman, D. (2004). "A Domain Interaction Map Based on Phylogenetic Profiling." *J. Mol. Biol.* 344 (5), 1331–1346. doi: 10.1016/j.jmb.2004.10.019
- Pavlova, S. L., Jin, L., Gasparovich, S. R., and Tao, L. (2013). "Multiple Alcohol Dehydrogenases But No Functional Acetaldehyde Dehydrogenase Causing Excessive Acetaldehyde Production From Ethanol by Oral *Streptococci*." *Microbiol. (Reading)* 159 (Pt 7), 1437–1446. doi: 10.1099/mic.0.066258-0
- Pearson, W. R. (2013). An Introduction to Sequence Similarity ("Homology") Searching. *Curr. Protoc. Bioinf. Chapter 3:Unit3.1*. doi: 10.1002/0471250953.bi0301s42
- Pushkar, S., Ji, X., Li, Y., Estilo, C., Yegnanarayana, R., Singh, B., et al. (2012). "Comparison of Oral Microbiota in Tumor and Non-Tumor Tissues of Patients With Oral Squamous Cell Carcinoma." *BMC Microbiol.* 12, 144. doi: 10.1186/1471-2180-12-144
- Qi, Y., Tastan, O., Carbonell, J. G., Klein-Seetharaman, J., and Weston, J. (2010). "Semi-Supervised Multi-Task Learning for Predicting Interactions Between HIV-1 and Human Proteins." *Bioinformatics* 26 (18), i645–i652. doi: 10.1093/bioinformatics/btq394
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). "Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies." *Nucleic Acids Res.* 43 (7), e47. doi: 10.1093/nar/gkv007
- Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). "Edger: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data." *Bioinformatics* 26 (1), 139–140. doi: 10.1093/bioinformatics/btp616
- Rost, B. (1999). "Twilight Zone of Protein Sequence Alignments." *Protein Eng.* 12 (2), 85–94. doi: 10.1093/protein/12.2.85
- Sean, D., and Meltzer, P. S. (2007). "GEOquery: A Bridge Between the Gene Expression Omnibus (GEO) and BioConductor." *Bioinformatics* 23 (14), 1846–1847. doi: 10.1093/bioinformatics/btm254
- Snoek-van Beurden, P. A., and Von den Hoff, J. W. (2005). "Zymographic Techniques for the Analysis of Matrix Metalloproteinases and Their Inhibitors." *Biotechniques* 38 (1), 73–83. doi: 10.2144/05381rv01
- Suzuki, R., Matsuno, S., Sakagami, H., Okada, Y., and Shirataki, Y. (2014). "Search of New Cytotoxic Crude Materials Against Human Oral Squamous Cell Carcinoma Using 1H NMR-Based Metabolomics." *Anticancer Res.* 34 (8), 4117–4120.
- Tamura, K., Stecher, G., and Kumar, S. (2021). "MEGA11: Molecular Evolutionary Genetics Analysis Version 11." *Mol. Biol. Evol.* 38 (7), 3022–3027. doi: 10.1093/molbev/msab120
- The Gene Ontology, Consortium (2019). The Gene Ontology Resource: 20 Years and Still GOing Strong. *Nucleic Acids Res.* 47 (D1), D330–D338. doi: 10.1093/nar/gky1055
- Vincent-Chong, V. K., Salahourifar, I., Karen-Ng, L. P., Siow, M. Y., Kallarakkal, T. G., Ramanathan, A., et al. (2014). "Overexpression of MMP13 Is Associated With Clinical Outcomes and Poor Prognosis in Oral Squamous Cell Carcinoma." *ScientificWorldJournal* 2014, 897523. doi: 10.1155/2014/897523
- Vogelmann, R., and Amieva, M. R. (2007). "The Role of Bacterial Pathogens in Cancer." *Curr. Opin. Microbiol.* 10 (1), 76–81. doi: 10.1016/j.mib.2006.12.004
- Wang, J., Sun, F., Lin, X., Li, Z., Mao, X., and Jiang, C. (2018). "Cytotoxic T Cell Responses to *Streptococcus* are Associated With Improved Prognosis of Oral Squamous Cell Carcinoma." *Exp. Cell Res.* 362 (1), 203–208. doi: 10.1016/j.yexcr.2017.11.018
- Wickham, H. (2011). "Ggplot2." *Wiley Interdiscip. Rev.: Comput. Stat* 3 (2), 180–185. doi: 10.1002/wics.147
- Wojcik, J., and Schächter, V. (2001). "Protein-Protein Interaction Map Inference Using Interacting Domain Profile Pairs." *Bioinformatics* 17 (Suppl 1), S296–S305. doi: 10.1093/bioinformatics/17.suppl_1.s296
- Yang, B., Dong, K., Guo, P., Guo, P., Jie, G., Zhang, G., et al. (2020). "Identification of Key Biomarkers and Potential Molecular Mechanisms in Oral Squamous Cell Carcinoma by Bioinformatics Analysis." *J. Comput. Biol.* 27 (1), 40–54. doi: 10.1089/cmb.2019.0211
- Yang, G., Peng, M., Tian, X., and Dong, S. (2017). "Molecular Ecological Network Analysis Reveals the Effects of Probiotics and Florfenicol on Intestinal Microbiota Homeostasis: An Example of Sea Cucumber." *Sci. Rep.* 7 (1), 4778. doi: 10.1038/s41598-017-05312-1
- Yen, C. Y., Chen, C. H., Chang, C. H., Tseng, H. F., Liu, S. Y., Chuang, L. Y., et al. (2009). "Matrix Metalloproteinases (MMP) 1 and MMP10 But Not MMP12 are Potential Oral Cancer Markers." *Biomarkers* 14 (4), 244–249. doi: 10.1080/13547500902829375
- Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). "ClusterProfiler: An R Package for Comparing Biological Themes Among Gene Clusters." *Omics* 16 (5), 284–287. doi: 10.1089/omi.2011.0118

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Fang, Yang and Liu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.