Assessment of Bioleaching Microbial Community Structure and Function Based on Next-Generation Sequencing Technologies

Shuang Zhou 1, Min Gan 2,*, Jianyu Zhu 2,*, Xinxing Liu 2 and Guanzhou Qiu 2

1 School of Public Health, Changsha Medical University, Changsha 410219, China; longzej@sina.cn
2 Key Laboratory of Biohydrometallurgy of Ministry of Education, School of Minerals Processing and Bioengineering, Central South University, Changsha 410083, China; lpdouzi@163.com (X.L.); zhongnanchw@163.com (G.Q.)
* Correspondence: ganmin0803@sina.com (M.G.); zhujy@csu.edu.cn (J.Z.)

Received: 4 September 2018; Accepted: 30 November 2018; Published: 17 December 2018

Abstract: It is widely known that bioleaching microorganisms have to cope with the complex extreme environment in which microbial ecology relating to community structure and function varies across environmental types. However, analyses of microbial ecology of bioleaching bacteria is still a challenge. To address this challenge, numerous technologies have been developed. In recent years, high-throughput sequencing technologies enabling comprehensive sequencing analysis of cellular RNA and DNA within the reach of most laboratories have been added to the toolbox of microbial ecology. The next-generation sequencing technology allowing processing DNA sequences can produce available draft genomic sequences of more bioleaching bacteria, which provides the opportunity to predict models of genetic and metabolic potential of bioleaching bacteria and ultimately deepens our understanding of bioleaching microorganism. High-throughput sequencing that focuses on targeted phylogenetic marker 16S rRNA has been effectively applied to characterize the community diversity in an ore leaching environment. RNA-seq, another application of high-throughput sequencing to profile RNA, can be for both mapping and quantifying transcriptome and has demonstrated a high efficiency in quantifying the changing expression level of each transcript under different conditions. It has been demonstrated as a powerful tool for dissecting the relationship between genotype and phenotype, leading to interpreting functional elements of the genome and revealing molecular mechanisms of adaption. This review aims to describe the high-throughput sequencing approach for bioleaching environmental microorganisms, particularly focusing on its application associated with challenges.

Keywords: high-throughput sequencing technology; RNA-seq; acidophiles; bioleaching; microbial community structure and function

1. Introduction

In the last decade, biomining-related bacteria as participators in the bioleaching processes have been intensively studied due to their importance in applications in the metal extraction from minerals. Leaching systems are considered a typical extreme environment, as they are often highly acidic (typically pH < 3) and usually contain increasing concentrations of iron, zinc, copper, and various other heavy metals [1]. Particularly, during bioleaching of mineral concentrates, heavy metals accumulate in the leaching solution. Metals are the metabolic requirements for microorganisms when they maintain the proper concentrations, but beyond certain concentrations they become toxic to the microorganism, mainly as a result of their ability to denature protein molecules [2]. However,
bioleaching microorganisms can better adapt to the most inhospitable environment. They play key roles as sulfur and/or iron oxidizers to efficiently enhance the dissolution of low-grade minerals in bioleaching systems (Figure 1A1). In addition, bioleaching is a complex process concerning the relationship of microbes with environmental factors (Figure 1A2) and the interaction between bioleaching microorganisms (Figure 1B) [3]. At present, the construction of acidophiles community and controlling bioleaching conditions have been piloted and demonstrated to accelerate dissolution and researchers continue to make progress in the mechanism studies for acidophilic microorganisms to solubilize ores [4–6]. Understanding the structure, functions, activities, and dynamics of microbial communities in bioleaching environments is important for the purpose of improving bioleaching rates [7–9].

**Figure 1.** (A1). Model for contact leaching catalyzed by biomining-related bacteria playing key roles as sulfur and/or iron oxidizers to enhance the dissolution of minerals. (A2). Proposed schematic diagram of interactions between substrates, abiotic drivers, biodiversity and ecosystem functions in a bioleaching system. (B). Concept model of roles of microorganisms involved in biogeochemical Fe & S cycling with C & N fixation/cycling and their interaction in bioleaching system. Reproduced with permission from Pablo Cardenas et al. [3], published by Elsevier, 2016.
A huge microbial diversity with wide metabolic potential that is influenced both by interactions with other bacteria and with the variable environment exists in most bioleaching systems [7,10]. To elucidate the functional response of microbial communities to changing environmental conditions has been challenging [11,12]. To address this challenge, numerous technologies have been developed (Figure 2). Cultivation-independent genomic approaches have significantly promoted our understanding of ecology and diversity of microbial communities in the environment. Function genes and 16S rRNA based molecular technologies—including fluorescence in situ hybridization (FISH), denaturing gradient gel electrophoresis (DGGE), quantitative real-time polymerase chain reaction (qRT-PCR), stable isotope probing (SIP) and related technologies (nanoscale secondary ion mass spectrometry, NanoSIMS), microarray and proteomics—have been developed to analyze the microbial community structure and gene diversities in various environments.

Stable isotope probing (SIP) has been used as probes or tracers to study dynamic processes/mechanisms in complex biological systems. It partly enhances our understanding of how individual microbial taxa affect ecosystem processes like element cycling by analyzing microbial diversity of intact assemblages. However, it is a qualitative technique capable of identifying some of the organisms that utilize a substrate, not a quantitative one capable of exploring the full range of variation in isotope incorporation among microbial taxa [13]. NanoSIMS in combination with stable isotope probing was applied to analyze and image biological samples, which helps us better understand biological processes happening in complex systems. However, compared with high-resolution microscopy techniques—such as scanning electron microscopy (SEM), transmission electron microscopy (TEM), and atomic force microscopy (AFM)—NanoSIMS do not reveal either detailed surface structures or subcellular structures. Thus some topographical or morphological information may not be gained for specific biological questions [14]. Proteomics to characterize proteins differentially expressed by various types of cell or cells subjected to different environmental conditions is an important tool to understand microbe mineral interaction and characterization of microbial biodiversity. In leaching processes, iron oxidation and sulfur reduction by bioleaching microbes occur mainly in the extracellular space. In agreement with this, several proteomics studies revealed protein-associated molecules present in the extracellular polymeric substance (EPS) layers that are able to accumulate sulfur and enhance the bioleaching of metal sulphides [15]. Researchers determined the differential response in the proteome of the acidophilic halophile, *Acidihalobacter prosperus* DSM 14174 (strain V6) at low and high chloride ion level, to thus understand the mechanism of tolerance to high chloride ion stress in the presence of low pH [16]. Through protein identification, stressing factors during chalcopyrite biomining were elucidated and new light was shed on resistance systems deployed by *Leptospirillum ferriphilum* [17]. Though proteomics provides direct information of the dynamic protein expression, giving us a global analysis, it should be combined with genomics and bioinformatics to systematically analyze all expressed cellular components so that a comprehensive picture of biology can be possibly grasped. Those methods mentioned above are useful for less diverse communities to some extent microbial diversity and couple microbial taxonomy diversity with diversified functions may not be reflected integrally due to low throughput.
Figure 2. Methods to discover and characterize microbial diversity and function.

Since the dawn of genetics, our view of the extent and complexity of microbe has been altered [18]. Increased gene-based tools have made it possible for researchers to study natural microbial communities’ structure and gene expression profiles through analysis of nucleic acids directly extracted from environmental samples [19,20]. DNA microarrays have provided scientists with the capability to simultaneously investigate thousands of fragments in a single experiment. The overwhelming wealth of knowledge generated by microarrays has created entirely new fields of research [21]. Over the years, hybridization-based microarray technologies as the dominant approaches have been instrumental in exploring gene expression. Proven outcomes of hybridization-based microarray approaches have accurately allowed deduction and quantification of the transcriptome [22–24]. Hybridization-based methods are typically dependent on incubation of fluorescently labeled cDNA with probes fixed onto solid surfaces (custom-made microarrays or commercial high-density oligo microarrays) [21]. Updated microarrays, for instance, tiling microarrays with probes representing the genome at a high density, can be used to map transcribed regions at a relatively high resolution and uncover novel transcripts. Ever since its first utilization in 1995 [25], microarrays have been widely used in transcriptomics by providing a high flux and relatively inexpensive access to genome-scale information, other than tiling arrays that interrogate genomes at high resolution. Nevertheless, microarray technology is generated with some inherent limitations [26–29], which include the dependence on preexisting knowledge about genome sequence, high levels of background noise as a result of cross hybridization, saturation of signals for high-abundant transcripts, and a narrow dynamic range of evaluating gene expression levels. Additionally, it is difficult to compare expression levels of different tests and sophisticated methods of normalization are needed.
Equally revolutionary technologies are currently emerging in the form of new methods of sequencing, termed massively parallel sequencing (MPS, also called next-generation/high-throughput sequencing) [30,31]. The intrinsic problems characterized with microarray methods were conquered with the introduction of high-throughput DNA sequencing technologies, which opened up new horizons for our understanding of bacterial gene expression and regulation by allowing RNA analysis through cDNA sequencing on a large scale [32]. DNA-based high-throughput sequencing metagenomics have been applied to reveal microbial communities in marine water [33], soil [34], activated sludge [35], human and animal guts [36,37], and animal waste [38]. However, questions of how natural bacterial assemblages respond to perturbations in environmental conditions are better answered by analysis of community mRNA than genomic DNA. RNA-Seq that directly sequences the cDNA is not limited to detecting the transcripts that accord with known genomic sequence. Therefore, identification, characterization, and quantification of new splice variants are allowed by RNA-Seq [39]. Additionally, RNA-Seq approach possesses other advantages over microarray technology, including low background signal, the inexistence of the ceiling for quantification and thus a much larger dynamic range of expression levels over which transcripts can be detected [40]. In the last few years, high-throughput RNA sequencing technologies have been added to the toolbox of microbial ecology and used to characterize the functional response of microbial communities to changing environmental conditions [41–43]. This approach allows the determination of the most highly transcribed genes of a community, thus providing first insights into community function under a specific set of environmental parameters.

Recently, the active development of next-generation sequencing (NGS) technology-based sequencing approaches has enabled comprehensive sequencing analysis of cellular RNA and DNA within the reach of most laboratories. The goals of this review are (1) to briefly introduce next-generation sequencing technologies; (2) to present the adoption of RNA-seq approach for complete genome, microbial community, and transcriptomes characterization; and (3) to describe some challenges confronted with sequencing technologies, and analyze the perspectives in light of rapid evolution of sequencing technologies.

2. NGS for Addressing the Challenge of Analyzing the Microbial Ecology in Bioleaching Environments

This review does not intend to describe sequencing technologies in depth, due to the pending publication of extensive outstanding reviews [30,31,44]. High-throughput sequencing methods were mainly based on 454 GS FLX (Roche, Basel, Switzerland), Genome Analyzer II (Illumina, San Diego, CA, USA) and SOLiD (Applied Biosystems, Foster City, CA, USA) platforms (Figure 3). Regardless of choosing sequencing platforms to address biological questions of interest, the disarmingly simple principle behind these sequencing methods is that to learn the content of a complex RNA/DNA sample, one can just sequence it directly without bacterial cloning as a prerequisite. Sequence census assays that use next-generation sequencing technologies were mainly applied for determining the sequence content and abundance of mRNAs, noncoding RNAs and small RNAs (RNA-seq) and for scanning whole genome profiles of chromatin immunoprecipitation (ChIP-seq), methylation sites (methyl-seq), and DNase I hypersensitivity sites (DNase-seq) [43].

RNA-Seq is another application of high-throughput sequencing and developed in multiple laboratories. RNA-seq, also called whole transcriptome sequencing, utilizes next-generation sequencing (NGS) technologies to profile RNA through sequencing cDNA, which is the conversion of isolated transcripts of interest. The microbial RNA-seq method involves several basic steps (Figure 3). The starting point is the extraction of RNA samples, followed by optional depletion of tRNA and rRNA, construction of cDNA libraries, sequencing on a selected massively parallel deep sequencing platform and the subsequent bioinformatic analysis of cDNA sequencing read histograms [45]. Over the past few years, this deep-sequencing-based approach has been exploited to reveal comprehensive insights to eukaryotic transcriptomes from yeast [46,47] to human [48,49] at
an unprecedented level. Recently, RNA-sequencing technology has been emerging as a developed tool for studying bacterial transcriptomes [40,50], and it has demonstrated high sensitivity for genes expressed either at low or very high levels, thus having a much large dynamic range, and accuracy in transcriptomes quantification and quantization [41,42,50]. In addition, the RNA-Seq technology permits the delineation of operons and untranslated regions, allowing the improvement and extension of sequence annotation [51], and the mapping of sequence data is more precise. This allows transcription to be studied at higher resolution by sequencing, also defining at single nucleotide resolution the transcriptional boundaries of genes and the expressed single nucleotide polymorphisms (SNPs) [41–43], thereby also permitting the study of more repetitive regions of the genome. Additionally, structural information can be used to refine annotated gene structures or propose novel gene models [51]. Other advantages of RNA-Seq compared to microarrays are that RNA-Seq data also show high levels of reproducibility for both technical and biological replicates. Generally, for gene expression analysis, RNA-seq is an advanced alternative solution to microarrays [52,53].

![Flow diagram of the steps involved in the genetic sequencing based on high-throughput sequencing platforms.](image)

**Figure 3.** Flow diagram of the steps involved in the genetic sequencing based on high-throughput sequencing platforms.

### 2.1. NGS for Genome Analysis

Next-generation DNA sequencing is dramatically accelerating biological insight to microbial life in many environments. Herein, we highlight progress in genomics of microbes from heap leaching conditions and related acidic mining environments. For better understanding the ecology of more complex natural environments, microbial ecology studies in model ecosystems are necessary. Acid mine-related environments have been identified as a model ecosystem, partially on account of biotic community characteristics in typical extreme environment [54], and it has been researched broadly because of their importance in application in the biomining industry [55,56]. In leaching systems, biochemical reactions with the participation of leaching microorganisms, coupled with chemical reactions, lead to the sulfide mineral dissolution and consequent metal release [57].
Also noteworthy is the effect of microorganisms to mineral bioleaching. The microbiology of leaching environments—including physiology of the most common community members, microbial successions [7,58], the relationship of the population dynamics with environmental factors [59,60], and the influence of community composition on ecosystem functioning [61,62]—have always been the target for research on bioleaching processes and mechanisms. Next generation sequencing technology enables the comprehensive analysis of genomes, which can recover information about their general characteristics, especially their metabolic potential [3,63].

By March 2016, 157 genomes of acidophiles were included in public databases. Among them, 29 (20%) are from microorganisms in bioleaching heaps or closely related mining environments [3]. These genomes are listed in Table 1. Additionally, there is plenty of relevant research on the genomics and metagenomics of acidophilic microorganisms from bioleaching heaps or related biomining environments. Through genomic analysis, genetic and predictive metabolic models of some microorganisms and the suggestion of ecophysiological interactions during bioleaching were produced [3].

### Table 1. Available genomes of acidophiles associated with bioleaching heaps or related biomining environments. Reproduced with permission from Pablo Cardenas et al. [3], published by Elsevier, 2016.

<table>
<thead>
<tr>
<th>Organism</th>
<th>NCBI Accession</th>
<th>Source</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Acidiphilium cupripusilum</em> BH2</td>
<td>LKH0000000000</td>
<td>Mineral sulfide ore, Myanmar</td>
<td>not available</td>
</tr>
<tr>
<td><em>Acidiphilium cupripusilum</em> JCM 13668</td>
<td>BBDK0000000000</td>
<td>Industrial-scale chalcocite bioleach heap, Myanmar</td>
<td>not available</td>
</tr>
<tr>
<td><em>Acidiphilium</em> sp. MBA-1</td>
<td>JYHS0000000000</td>
<td>Bioremediation bioreactor pulp, Russia</td>
<td>not available</td>
</tr>
<tr>
<td><em>Sulfobacillus acidocaldarius</em> Ron12/1</td>
<td>NC_000247</td>
<td>Uranium mine heaps, Germany</td>
<td>[64]</td>
</tr>
<tr>
<td><em>Acidiphilium</em> angustum ATCC 35903T</td>
<td>JNH0000000000</td>
<td>Waste coal mine waters, USA</td>
<td>not available</td>
</tr>
<tr>
<td><em>Acidiphilium</em> cryptum JP-5</td>
<td>NC_000484</td>
<td>Acidic coal mine lake sediment, Germany</td>
<td>not available</td>
</tr>
<tr>
<td><em>Acidithiobacillus caldus</em> ATCC 51756T</td>
<td>CP000916</td>
<td>Coal spoil enrichment culture, UK</td>
<td>[66]</td>
</tr>
<tr>
<td><em>Acidithiobacillus caldus</em> SM-1</td>
<td>NC_015350</td>
<td>Pilot bioremediation reactor, China</td>
<td>[67]</td>
</tr>
<tr>
<td><em>Acidithiobacillus ferrivorans</em> CF27</td>
<td>CCCC0000000000</td>
<td>Abandoned copper/cobalt mine drainage, USA</td>
<td>[68]</td>
</tr>
<tr>
<td><em>Acidithiobacillus ferrivorans</em> SS3</td>
<td>NC_015942</td>
<td>Enrichment culture from mine-impacted soil samples, Russia</td>
<td>[69]</td>
</tr>
<tr>
<td><em>Acidithiobacillus ferrivorans</em> ATCC 2327T</td>
<td>NC_011761</td>
<td>Acid, bituminous coal mine effluent, USA</td>
<td>[70]</td>
</tr>
<tr>
<td><em>Acidithiobacillus ferrivorans</em> ATCC 53993</td>
<td>NC_011206</td>
<td>Copper deposits, Armenia</td>
<td>not available</td>
</tr>
<tr>
<td><em>Acidithiobacillus</em> sp. GG-3-21</td>
<td>AEB0000000000</td>
<td>Mine water, India</td>
<td>not available</td>
</tr>
<tr>
<td><em>Acidithiobacillus thioparus</em> A01</td>
<td>A2M0000000000</td>
<td>Wastewater of coal dump, China</td>
<td>[71]</td>
</tr>
<tr>
<td><em>Acidithiobacillus thioparus</em> ATCC 1937T</td>
<td>AFOH0000000000</td>
<td>Kinneredg clay, UK</td>
<td>[72]</td>
</tr>
<tr>
<td><em>Acidithiobacillus thioparus</em> Licanaturant</td>
<td>JMEB0000000000</td>
<td>Copper mine, Chile</td>
<td>[73]</td>
</tr>
<tr>
<td><em>Acidithrix ferrivorans</em> DSM 28176T</td>
<td>JXY0000000000</td>
<td>Acidic stream drainage in abandoned copper mine, UK</td>
<td>[74]</td>
</tr>
<tr>
<td><em>Ferrimicrobium acidiphilum</em> DSM 1949T</td>
<td>JQKF0000000000</td>
<td>Mine water, UK</td>
<td>[75]</td>
</tr>
<tr>
<td><em>Leptospirillum ferrooxidum</em> DSM 14447T</td>
<td>JPG0000000000</td>
<td>Enrichment culture, Peru</td>
<td>[76]</td>
</tr>
<tr>
<td><em>Leptospirillum</em> sp. Sp-Ci</td>
<td>LGSH0000000000</td>
<td>Industrial bioremediation solution, Chile</td>
<td>[77]</td>
</tr>
<tr>
<td>“Ferrovum myosaficiens” P3G¹</td>
<td>JIPO0000000000</td>
<td>Stream draining an abandoned copper mine, UK</td>
<td>[78]</td>
</tr>
<tr>
<td><em>Ferrovum</em> sp. JA12</td>
<td>LJW0000000000</td>
<td>Pilot treatment plant water, Germany</td>
<td>[79]</td>
</tr>
<tr>
<td><em>Ferrovum</em> sp. Z-31</td>
<td>LRRD0000000000</td>
<td>Acid mine drainage water, Germany</td>
<td>not available</td>
</tr>
<tr>
<td><em>Ferrovum</em> sp. PN-318</td>
<td>LQA0000000000</td>
<td>Acid mine drainage water, Germany</td>
<td>not available</td>
</tr>
<tr>
<td>“<em>Acidithiobacillus</em> ferrivorans” DSM 5130T</td>
<td>LPVF0000000000</td>
<td>Neutral drainage from copper mine, Brazil</td>
<td>[80]</td>
</tr>
<tr>
<td><em>Sulfobacillus acidiphilus</em> DSM 10332²</td>
<td>NC_016684</td>
<td>Coal spoil heap, UK</td>
<td>[81]</td>
</tr>
<tr>
<td><em>Sulfobacillus thermosulfidoxidans</em> CBAR13</td>
<td>LGRG0000000000</td>
<td>Percolate solution of a bioremediation heap in copper mine, Chile</td>
<td>not available</td>
</tr>
<tr>
<td><em>Sulfobacillus thermosulfidoxidans</em> Cutypay</td>
<td>ALW0000000000</td>
<td>Naturally mining environment, Chile</td>
<td>[82]</td>
</tr>
<tr>
<td><em>Sulfobacillus thermosulfidoxidans</em> DSM 5923T</td>
<td>(250210005)²</td>
<td>Spontaneously heated ore deposit, Kazakhstan</td>
<td>not available</td>
</tr>
<tr>
<td>Bioremediation heap surface Metagenome</td>
<td>(4645533.3)²</td>
<td>Dexing Copper Mine, China</td>
<td>[9]</td>
</tr>
<tr>
<td>Bioremediation heap sample Metagenome</td>
<td>(4554863.3)²</td>
<td>Dexing Copper Mine, China</td>
<td>[83]</td>
</tr>
<tr>
<td>Bioremediation heap sample Metagenome</td>
<td>(4554867.3)²</td>
<td>Dexing Copper Mine, China</td>
<td>[83]</td>
</tr>
<tr>
<td><em>Acidithiobacillus thioparus</em> CLST</td>
<td>NZ_LGYM00000020.1</td>
<td>Gorbea salt flat, northern Chile.</td>
<td>[84]</td>
</tr>
</tbody>
</table>

Note. T = type strain; * sequence only available in IMG-JGI where the IMG Taxon ID value is provided; # sequence only available in MG-RAST where the correspondent ID value is provided.

*Acidithiobacillus ferrivorans* (*A. ferrivorans*), chemolithoautotrophic bacteria can obtain energy from oxidation of elemental sulfur and ferrous compounds to maintain cell growth. The gammaproteobacterium *A. ferrivorans* is adapted to growth in the extreme environment and accounts for a considerable part in mine-related contexts. It is commonly recognized as a model
organism for the investigation of metal sulfide bioleaching [70]. The contribution of *A. ferrooxidans* to mineral bioleaching has been widely studied, whereas to gain insight into their biology, bioinformatic analysis of genome information has been a major route.

Over a decade ago, the genome of *A. ferrooxidans* ATCC 23270 was sequenced and first published in draft form [85]. Based on analysis of microbial genomes, reconstruction of amino acid metabolism and sulfur assimilation [86,87], prediction of fur regulation [88], acyl homoserine lactone production [89], quorum sensing [90,91], the formation of extracellular polysaccharide [92,93], carbon metabolism and iron and sulfur oxidation [94–96] were carried out. Furthermore, predictive models of genetic and metabolic potential of bioleaching bacteria were solidified and extended [3], ascribing to the published complete genome sequence of *A. ferrooxidans* in 2008. Generally, the first glimpse of genome of *A. ferrooxidans* by sequencing accelerates our understanding of acidophilic life in bioleaching conditions. However, this information is insufficient to allow a reasonable description of the genetic complexity and the prediction of metabolic capabilities and interactions with other acidophiles in bioleaching processes. Therefore, the genomes from *A. ferrooxidans* as model organism cannot serve as substitutes for constructing genetic and metabolic models of another bioleaching bacterium [3]. Additionally, there are other microorganisms involved in bioleaching in addition to *A. ferrooxidans*. In order to know the linkage between ecophysiological interactions with ecological functions, it is urgent to study the nucleotide sequences of another various bacterium. Ever since 2008, the implementation of many genome sequencing projects on strains has made the draft genomes of other bioleaching bacteria become exploitable.

*Leptospirillum ferriphilum* (*L. ferriphilum*), chemolithoautotrophic, and acidophilic bacteria, can get energy through Fe$^{2+}$ oxidation, and they are one of key players in the sulfide mineral bioleaching system due to their capability of iron oxidization. Four subspecies of *Leptospirillum* including *Leptospirillum ferrooxidans* (*L. ferrooxidans*), *Leptospirillum rubarum* (*L. rubarum*), *Leptospirillum ferrodiazotrophum* (*L. ferrodiazotrophum*), and *L. ferriphilum*, have been identified. However, limited knowledge of ways to obtain energy and nutrients for growth, mechanisms for nitrogen fixation, adsorption of bacteria to mineral surface, and the ability to adapt bioleaching conditions with acidic pH, high metal concentrations and reactive oxygen species, hinder the understanding of *Leptospirillum* that lagged by comparison to the *Acidithiobacillus* genus. The illumination of metabolic properties and ecophysiological interactions in leaching systems was blocked, ascribing to the only available draft genome of the *L. ferriphilum* strain in spite of already published genomes of *L. ferriphilum* strains [17]. Stephan Christel centered on in-depth analysis of characterization of this organism’s metabolic potential via sequencing the *L. ferriphilum*T DNA and reconstructed the model of the genomic potential observed in the *L. ferriphilum*T genome [17]. The genetic information provided by this study advanced the investigation of the role of *L. ferriphilum*T in the acid mine and bioleaching processes.

The bacteria leaching of sulphide minerals is a process that needs the involvement of both iron- and sulfur-oxidizing microorganisms. *Acidithiobacillus thiooxidans* (*A. thiooxidans*), sulfur oxidizer, gains energy through oxidizing elemental sulfur (S$^{0}$) and sulfur compounds to support cell growth and carry out bioleaching processes [95]. Inadequate published data on *A. thiooxidans* genome limited our study of its physiology [97]. The advent of NGS allows for sequencing the *A. thiooxidans* whole genome, and consequently the construction of a preliminary model of its whole genome. Especially, the genomic elements related to sulfur oxidation were studied. All these findings accelerated the understanding of its bioleaching potential and adaptive capacity to ore leaching environment. Three genome sequences of *A. thiooxidans* ATCC 19377, A01, and CLST have been published in draft form, which provided valuable information on general features of *A. thiooxidans* [71,72,84]. In order to acquire new insights to the bioleaching characteristics of *A. thiooxidans*, Dante Travisany [73] conducted a gene study on *A. thiooxidans* strain isolated from a Chilean copper mine, and in 2014, a new genome sequence from Licanantay (DSM17318) was released by them. By genetic comparison analysis with *A. thiooxidans* ATCC 19377 and A01, a certain similarity in coding sequences appeared in *A. thiooxidans* Licanantay.
Additionally, the unique genes observed in the genome of *A. thiooxidans Licanantay* suggests its adaptation to specific extreme environment and its bioleaching potential.

In general, NGS technology allowing processing DNA sequences can produce draft genomic sequences of more bioleaching bacteria, which provides an opportunity to predict models of genetic and metabolic potential of bioleaching bacteria and ultimately deepens our understanding of bioleaching microorganisms.

### 2.2. NGS for Analysis of Bacterial Diversity Present in the Ore Leaching Environment

Bioleaching microorganisms inhabiting extreme environment are involved in the biochemical cycling of elements, such as sulfur, iron, and various metals. They play integral and unique roles in leaching systems, and their structure, interaction, and dynamics to leaching conditions are critical to mineral dissolution and metal recovery. Gaining insight to microbial community structure and functions is critical for understanding the bioleaching process and eventually improving leaching efficiency. To investigate the community structure, searching for more available molecular markers and techniques has always been a subject of importance. The appropriate molecular marker used for microbial phylogenetic reconstruction, identification, and classification of strain is the 16S ribosomal RNA, partially due to its strain specialty and highly conserved sequence and structure [98]. NGS with traits of high throughput, specificity, and relative quantification easily detect more microbial diversity. Effective high-throughput sequencing that focuses on targeted phylogenetic markers (e.g., 16S rRNA) [99–101] has been applied to characterize community diversity.

To date, some studies have used this kind of approach to assess the dynamics of bioleaching microorganisms inhabiting industrial or natural environment. Baker and colleagues [102] applied metagenomics analyses of acidophilic communities in acid mine drainage (AMD) at Iron Mountain California, expanding our view from individual genes and cultures to entire communities. Additionally, metagenome-scale analysis of bioleaching heaps [9,101,103] and acidic hot springs [104] yield insights into the structure and function of microbial communities, allowing the establishment of correlations between the occurrence of certain microbes, their activities and the geochemistry of cognate sites [105]. With the 16S rRNA gene sequencing, the shift of microbial communities in leaching heap, leaching solution (LS), and sediment subsystems in Dexing Copper Mine were examined by Jiaojiao Niu (Figure 4) [62], showing that *Acidithiobacillus, Leptospirillum, and Acidiferrobacter* (S and Fe oxidizers) were dominate strains in leaching heap and leaching solution while *Acidiphilium* (S and Fe reducer) were more abundant in the sediment. It indicated that the significant shift in community structures of subsystems might be a result of different geochemical conditions. NGS-based analyses for the microbial ecology within acidophilic communities in the Pb/Zn mine in China and low-temperature AMD waters originating from sulfide mine in Sweden have also be reported. The relative abundance of *ferrivorans*-like, *A. ferrooxidans*-like and *A. thiooxidans*-like strains has allowed for variability analyses [106]. All these found that the relative abundance of iron-oxidizing *Acidithiobacillus* species varies consistently with changing Fe$^{3+}$ and Cu$^{2+}$ concentrations [107], and it was dominant in the systems with lower ferric to ferrous iron concentrations and pHs above 3. However, sulfur-oxidizing *Acidithiobacillus* were dominant species in hot springs with rich sulfide.

In general, the NGS-based method with sufficient sequencing depth allows to capture the genomic information and ecological roles of low-abundance populations. It provides information concerning the dynamically shifted microbial communities to geochemical conditions. The advent of sequencing technologies studying the compositions and dynamics of microbial communities at the rRNA level has created unprecedented opportunities to reveal the ecology and evolution of extreme acidic microbial assemblages.
Figure 4. (A) Microbial community differentiation between the leaching heap (LH) and the pregnant leach solution (PLS) of the Dexing copper mine (a biological heap leaching system) using whole genome sequencing (WGS) metagenomic strategy and GenBank, RefSeq, and SEED databases. Reproduced with permission from Hu et al. [103], published by Elsevier, 2015. (B) Microbial community structure in LS, LH, and sediment systems and the shared and distinct OTUs of three subsystems showed by Venn diagram. Reproduced from Niu et al. [62], published by Springer, 2016.
2.3. NGS for Analysis of Gene Expression in Bioleaching Microorganisms

In extreme ore leaching environments, bioleaching microorganisms mainly have to maintain a near-neutral intracellular pH, preclude invasion of extraneous nucleic acid substances, respond to scarce availability of substrates and solvent extraction process, and resist to metal ions (Figure 5). It is of great importance to know how they thrive and develop in an extreme environment. Transcriptional analysis helps to fully understand biological processes in bioleaching microorganisms, such as development, adaptive evolution, and stress response. Unlike static genomes, transcripts dynamically change with developmental stage, physiological condition, and external environment. High-throughput mRNA sequencing technologies, termed RNA-seq, can be for both mapping and quantifying transcriptome and have demonstrated high efficiency in quantifying the changing expression level of each transcript under different conditions. They are now displacing microarrays and being exploited for transcriptional analysis as the preferred method. RNA-seq is a powerful tool for dissecting the relationship between genotype and phenotype, leading to interpreting functional elements of the genome and revealing the molecular mechanisms of adaption.

Figure 5. Proposed model for processes involved in adaptation of acidophiles to bioleaching heaps or related biomining environments. Those associated with oxidative stress response, low pH, and heavy metal resistance are shown. Reproduced from Peng et al. [108], published by Elsevier, 2017.

In order to expound adaptation mechanisms of bioleaching microorganisms to the extreme environment, information from genomic and transcriptomic assays is in demand. Stephan Christel [17] and colleagues used multi-omics to reveal the lifestyle of the acidophilic, mineral-oxidizing model species *L. ferriphilum*. Through RNA transcript sequencing and proteomics, the genes for growth using Fe$^{2+}$ as substrate and during chalcopyrite biomining were identified. According to their study, a previously unrevealed cluster for nitrogen fixation was captured, and metabolic processes including energy conservation, carbon dioxide fixation, pH homeostasis, metal resistance, oxidative stress management, chemotaxis and motility, quorum sensing and c-di-GMP, and biofilm formation were illuminated through analysis of mRNA transcripts. In addition, heavy metal resistance, chemotaxis,
and motility systems of \textit{L. ferriphilum}\textsuperscript{T} grown with chalcopyrite were found at higher expression levels in comparison with those in \textit{L. ferriphilum}\textsuperscript{T} grown with Fe\textsuperscript{2+} as substrate, which explained that elevated exposure of cells grown on minerals to heavy metals and rapid cells attachment to mineral surface. This study enhanced our understanding of the role of \textit{L. ferriphilum}\textsuperscript{T} in acid mine and rock drainage as well as bioleaching processes, and optimization bioleaching conditions for metal extraction.

Many studies on the adaptation mechanisms of bacteria to acid mine drainage have been reported. However, fewer have been carried out on microorganisms in acid mine drainage at high altitude. On the basis of transcript analysis using RNA-seq, Tangjian Peng \[108\] uncovered the adaptation mechanisms of \textit{A. ferrivorans} strain YL15 to the acid mine drainage environment in Yulong copper mine in Southwest China. Many genes of \textit{A. ferrivorans} strain YL15 residing in low-temperature condition that are involved in protein synthesis, transmembrane transport, energy metabolism and chemotaxis were found to show a higher expression level. Additionally, a bacterioferritin Dps (DNA binding proteins) gene had higher RNA transcript counts at low temperature, which was related to DNA protection against oxidative stress at low temperature. Through transcriptomic analysis, the cold adaptation mechanisms of \textit{A. ferrivorans} strain YL15 were illuminated, and a predictive model of the adaption of \textit{A. ferrivorans} strain YL15 to the alpine acid mine drainage environment was proposed. The valuable information from this study deepens our understanding of adaption mechanism of bioleaching strain.

There have been relevant studies describing the dynamic of the structure and function of the microbial community in bioleaching heaps \[109,110\]. Based on bioinformatics analyses of available genomes, a proposed preliminary model relates the dynamics with three different pathways of CO\textsubscript{2} fixation including Calvin Benson Bassham cycle (CBB), the reductive citric acid cycle (rTCA) and the 3-hydroxypropionate/4-Hydroxybutyrate cycle. However, it is hard to support this presumption due to lack of proteomic or transcriptomic evidence. Using RNA-seq, Sabrina Marín \[111\] studied carbon fixation pathways at the transcriptomic level in a controlled heap-like environment. Transcriptomic evidence showed that the active CBB and rTCA key genes were detected in the bioleaching environment, confirming the proposed active function of the regulation system in this bioleaching condition. These findings promote our understanding of the positive effect of high temperature on chalcopyrite leaching, thus optimizing bioleaching technology.

It is widely known that bioleaching microorganisms have to cope with complex extreme environment. Microbial ecology relates to community structure and function, and this varies across environmental types. However, analyses of microbial ecology of bioleaching bacteria are still a challenge. NGS technologies provide valuable insights into this aspect of gene expression profiling and therefore enhance our understanding of ecology and evolution of extreme acidic microbial assemblages.

3. Challenges and Prospect

The effectiveness of high-throughput sequencing as a tool for the identification of bioleaching microbial species and gene expression profiling has been demonstrated. However, it is still confronted with several challenges. First, experimental data can probably not reflect the actual composition of the sample due to bias introduction by cDNA libraries preparation. Several manipulation stages during the production of cDNA libraries include reverse transcription, ligation, and random priming. During reverse transcription, the first strand cDNA as well as the second strand are sometimes synthesized by enzymes. Inefficient or efficient RNA-RNA or RNA-DNA ligation at some sequences, combined with uneven coverage caused by random priming, may create different bias in the outcome \[112,113\]. All these can complicate the use of RNA-Seq in transcript profiling. Second, RNA-Seq faces informatics challenges resulting from the large amount of data. These data have to be processed for reconstruction of full transcripts, individual variant analysis, and even quantitation of expression levels for each transcript and gene, all of which should be assisted by a variety of software and bioinformatics tools and significant levels of expertise with programming skills. Thus, it is a challenge to analyze
large datasets produced by the different NGS technologies. Third, in order to definitely meet the needs of high-throughput sequencing work, the DNA or RNA extracted should be relatively high in concentration, and this indicates that the genetic analysis based on NGS may be limited because of small amount of some biological samples. Finally, higher cost for more sequencing depth which is required for adequate sequence coverage must be taken into consideration.

In spite of these challenges, high-throughput sequencing has already created a tremendous amount of influences on our understanding of microbial ecology of the leaching environment. It allows us to investigate the transcription at single-nucleotide resolution, which enriches our knowledge of microbial diversity, and will undoubtedly show us many different approaches adopted by bioleaching bacteria to solve problems encountered in their respective niches. As the sequencing technology develops rapidly and its cost decreases, high-throughput sequencing as culture-independent approach has opened up new avenues for genomes of complex microbial communities and gene expression, and it is taking place of microarrays as the preferred method for studying microbial communities and gene expression profiling, thus helping us understand evolutionary mechanisms and dynamics.

4. Conclusions

With the purpose of improving bioleaching rate, understanding the structure, functions, activities, and dynamics of microbial communities in bioleaching environments is always of concern. Next-generation sequencing technologies are dramatically accelerating biological insight to microbial life in these extreme conditions. Thus, this paper provides a review of describing the high-throughput sequencing approach, particularly focusing on its application associated with challenges in understanding bioleaching environmental microorganisms. NGS technology can process DNA sequences and can produce available draft genomic sequences of more bioleaching bacteria, which provides an opportunity to predict models of genetic and metabolic potential of bioleaching bacteria. Moreover, the NGS-based method studying the compositions and dynamics of microbial communities at the rRNA level, has created unprecedented opportunities to reveal the ecology and evolution of extreme acidic microbial assemblages. Additionally, it provides valuable insights into this aspect of gene expression profiling and therefore enhances our understanding of ecology and evolution of extreme acidic microbial assemblages. In conclusion, in spite of challenges, high-throughput sequencing has already had a tremendous influence on our understanding of the microbial ecology of leaching environments.

Author Contributions: S.Z. and M.G. designed, conducted, and wrote this whole review; G.Q. carried out English editing of the whole paper; J.Z. and X.L. collected references and materials for this paper.

Funding: This research was funded by the key project of Education Department of Hunan province (17A025); Opening Foundation of Key Laboratory of Biohydrometallurgy, Ministry of Education, Central South University, Changsha 410083, China (MOEKL1704); Hunan key laboratory cultivation base of the research and development of novel pharmaceutical preparations (2016TD1029); National Natural Science Foundation of China (51804350 and 41773089); postdoctoral foundation for MG from Chinese PD science foundation (2017M610506 and 2018T110842); PD research funding plan in Hunan and Central South University (18S690); the Key Discipline Project (Food Hygiene and Nutrition) of Changsha Medical University.

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMD</td>
<td>Acid Mine Drainage</td>
</tr>
<tr>
<td>CBB</td>
<td>Calvin–Benson–Bassham cycle</td>
</tr>
<tr>
<td>DGGE</td>
<td>Denaturing Gradient Gel Electrophoresis</td>
</tr>
<tr>
<td>FISH</td>
<td>Fluorescence in situ hybridization</td>
</tr>
<tr>
<td>LH</td>
<td>Leaching Heap</td>
</tr>
<tr>
<td>LS</td>
<td>Leaching Solution</td>
</tr>
<tr>
<td>MPS</td>
<td>Massively Parallel Sequencing</td>
</tr>
</tbody>
</table>
NanoSIMS  Nano-scale Secondary Ion Mass Spectrometry  
NGS  Next-generation sequencing  
PLS  Pregnant Leach Solution  
qRT-PCR  quantitative Real-Time Polymerase Chain Reaction  
rTCA  Reductive Citric Acid Cycle  
RICS  Reduced Inorganic Sulfur Compounds  
SIP  Stable Isotope Probing  
SNPs  Single Nucleotide Polymorphisms  
WGS  Whole Genome Sequencing

References


40. Haas, B.J.; Chin, M.; Nusbaum, C.; Birren, B.W.; Livny, J. How deep is deep enough for RNA-seq profiling of bacterial transcriptomes? *BMC Genom.* 2012, 13, 734. [CrossRef]


42. Zhou, X.; Ren, L.; Meng, Q.; Li, Y.; Yu, Y.; Yu, J. The next-generation sequencing technology and application. *Protein Cell* 2010, 1, 520–536. [CrossRef] [PubMed]


70. Valdes, J.; Pedroso, I.; Quatrini, R.; Dodson, R.J.; Tettelin, H.; Blake, R., II; Eisen, J.A.; Holmes, D.S. *Acidithiobacillus ferrooxidans* metabolism: From genome sequence to industrial applications. *BMC Genom.* 2008, 9, 597. [CrossRef]


100. Jia, Y.; Sun, H.; Chen, D.; Gao, H.; Ruan, R. Characterization of microbial community in industrial bioleaching heap of copper sulfide ore at monywa mine, myanmar. Hydrometallurgy 2016, 164, 355–361. [CrossRef]


