Using genomics in mine reclamation

L.H. Fraser  Natural Resource Sciences, Thompson Rivers University, Canada
H.W. Garris  Natural Resource Sciences, Thompson Rivers University, Canada
S.A. Baldwin  Chemical and Biological Engineering, University of British Columbia, Canada
J.D. Van Hamme  Biological Sciences, Thompson Rivers University, Canada
W.C. Gardner  Natural Resource Sciences, Thompson Rivers University, Canada

Abstract

For the majority of mines, closure succeeds when healthy, self-sustaining ecosystems develop on previously mined lands. In British Columbia, Canada, the regulations require reclamation of ecosystems; however, there are few specified targets, and those that are presented are vague. Genomics technologies may provide the key to both understanding the elements necessary to recreate functional ecosystems and provide sufficient benchmarks for success. In this review, we highlight the use of genomics to meet mine closure goals, enhance ecosystem development and optimise ecosystem services inherent in self-sustaining reclaimed ecosystems. We outline practical steps for applying genomics technologies to characterise the composition and activity of microbial communities in soils and treatment substrates. From this framework, we address the state of the science and how recently developed techniques have transferable value to mine reclamation. We then define three areas in which genomics technologies have already proven effective at informing management and reclamation of mine sites in the form of bioreactors, passive treatment systems and novel gene discovery. Finally, we speculate on the future applications of genomics technologies and the necessary steps to integrate these data into comprehensive management of mined sites.

1 Introduction

In an ideal system, mines operate with minimal environmental impact within and outside of the operational footprint. After the period of profitable extraction is reached, a desirable condition is recreated, be it forest, pasture or suburban neighbourhood, so that the biota (soil microbes, trees, and concerned citizens alike) can function within a natural, self-sustaining ecosystem. In reality, mine sites often leave a legacy, including perpetually altered plant communities (Holl, 2002); elevated contaminants in surface and groundwater (Cidu et al., 2001); thin, compact soils (Skousen et al., 2009); altered soil function (Mummey et al., 2002); and magnification of contaminants within the food chain (Allan, 1995; Muscatello and Janz, 2009).

In Canada, government regulations on mining operations set the standards for reclamation success, and as a result, direct the fate of mined lands. The British Columbia Mines Act and Health, Safety and Reclamation Code for Mines (the Code) specify that reclamation must satisfy the requirements of the chief inspector (Government of British Columbia, 2008). The legislation is vague and subject to interpretation (by both the mining companies and the inspectors), so there has been variability in goals and measures of success. Historically, goals were set for plant productivity, likely because productivity is relatively easy to measure and can be associated with ecosystem function (Britton, 1998). The current Code (10.7.5) expresses a goal for equivalent land capability, which is a challenge to achieve because there are no set predefined targets. There has been a recent trend to consider mine closure in terms of the whole ecosystem, with all of its functions and services.

Ecosystem services can be defined as “the aspects of ecosystems utilised (actively or passively) to produce human well-being” (Fisher et al., 2009). The ecosystem services for reclaimed mined lands may include forage...
for cattle and wildlife, timber, clean air and water, carbon sequestration, biodiversity and culturally significant natural products related to traditional practices and medicines. Canadian regulations exist to protect these services, or ensure their continuation, during and after mining activities take place. Despite pervasive efforts to meet and improve upon regulatory guidelines for ecosystem services at previously mined sites, there is a large degree of uncertainty in mine land restoration. Consistent application of revegetation, soil amendments and regrading treatments can lead to very different results, even on the same site (Martínez-Ruiz and Marrs, 2007). Why such inconsistency? Many authors attribute it to variability in starting conditions (the prevailing mineralogical substrate), slope, aspect and myriad unmeasured factors (Martínez-Ruiz and Marrs, 2007), with one emerging frequently — the soil microbiota (Harris, 2003).

Microorganisms are catalysts for soil formation. They are responsible for creating complex microenvironments that lead to nutrient uptake in plants, semi-homeostatic water and chemical regulations, and to overall resilience to erosion, contamination and invasion by exotic plant species. New soils that form on exposed bedrock often follow a predictable successional sequence. Microbial communities form on weathered inorganic substrates, fix atmospheric nitrogen and CO\textsubscript{2} and contribute the key elements of organic life. Multi-cellular plants and animals are relative latecomers to the developing community, arriving when soil formation processes have provided sufficient organic resources to sustain them (see Frouz et al., 2008, for a discussion of this process on previously mined sites). Chronologically, this natural approach can be considered bottom-up, with soil organisms and processes establishing and preparing the substrate before larger plants and animals arrive. The science of restoration has not developed a mechanism for recreating (much less accelerating) this bottom-up process. Rather, practitioners often attempt to recreate the desired ecosystem irrespective of the natural chronological sequence by superimposing soil amendments, and with seeding or planting plugged vascular plants (Tordoff et al., 2000). This practice is not undertaken out of ignorance of the role soil microbial communities play, but because tools to observe, quantify and manipulate these communities are not within the typical organisation’s toolkit.

Until recently, techniques for observing and quantifying microbiota have been extremely restrictive (Ficetola et al., 2008). Less than one per cent of all soil microbial organisms have been cultured in a laboratory setting (Hugenholtz et al., 1998; Harris, 2009). As one alternative, microbiologists have developed techniques for quantifying the mass-action of microbes in soils, including the respiration and production rates of metabolic products. Genomics technologies offer us an opportunity to observe the complexity of microbial communities as they form on mined sites, and to apply ecological theory to soil community formation and structure in a way that has until very recently been impossible. In recent decades, microbial genomics has been applied to mine sites, as will be discussed, but it has yet to be incorporated into a comprehensive monitoring and troubleshooting paradigm for mine closure and mine site restoration. In this paper, we provide a context for genomics in the mining community and highlight specific applications of genomics and metagenomics to mine closure.

2 Genomics techniques: From sample to management decision

2.1 Sampling

Microbial genomics begins with a source, either an individual organism or sample material comprising many organisms. Soil contains a combination of living, dead and dormant organisms, all of which contribute to the soil’s genetic signature. Representative sampling of soils on mine sites for their microbial community structures is a challenge, as mine sites are large, often incorporating entire landscapes with varying geophysical properties. Furthermore, soil communities are notoriously patchy in distribution (among habitat types and both horizontally and vertically within each soil stratum) (Foster, 1988; Ranjard and Richaume, 2001). Therefore, it is not feasible to sample every habitat type and soil stratum within a mined landscape. Rather, sampling should target habitats and soil strata that may provide signs of imminent degradation of water quality (streams and associated wetland areas within and adjacent to the mined land catchment), contaminant sequestration (bioreactor substrates, or organic-rich soils receiving runoff from the mine site) or reclamation success (processed parent material, organic amendments and unimpacted reference sites).
There are a few major concerns to consider when designing a sampling protocol for genomic analysis. First, the soil chemistry and water content of samples change rapidly upon collection, possibly leading to shifts in the activity and abundance of microbes within the soil, especially when collected from anoxic substrates. Biological activity must therefore be suspended as quickly as possible following collection, which generally involves flash freezing of samples with liquid nitrogen (−195°C) or on blocks of dry ice (−79°C) when liquid nitrogen is not available or is too dangerous to use (e.g. on a boat). Ethanol (95%) can be sprayed on blocks of dry ice to accelerate heat transfer. This freezing process suspends metabolic activity in the soil and preserves in situ bacterial abundance. Second, replication should be large enough to account for the high level of variation in soil microbial community composition and structure, with each replicate representing a homogeneous mixture of multiple samples (often soil cores) taken in the field. Biogeochemical gradients across physical interfaces and within sediments can be very steep. Therefore, preservation and assessment of layers is preferable (see de Gruijter et al., 2006, for a comprehensive discussion of sampling designs specific to natural resource monitoring). As microbial communities depend very strongly on local conditions (water/soil chemistry and biological interactions), sufficient metadata should be recorded along with each sample to enable this variation to be described for a particular community. As a minimum, this should include pH, temperature, dissolved oxygen, conductivity, nitrate, nitrite and phosphorus. Finally, because of their ability to bring oxygen and carbon resources into otherwise anaerobic soils, plant rooting zones can play a pivotal role in determining microbial community composition (Marschner et al., 2001; Marschner et al., 2004), and samples taken from vegetated substrates should account for the composition and characteristics of co-occurring plant communities.

2.2 Extraction

DNA extraction involves two steps: (1) breaking apart cell walls and membranes (cell lysis) through some combination of heating, sonication and/or chemical treatment, and (2) isolation/concentration of DNA via filtration (Picard et al., 1992; Zhou et al., 1996). Although conceptually similar, different lysis methods can produce conflicting results (Martin-Laurent et al., 2001; Carrigg et al., 2007).

There are three major challenges to consistency in the extraction process. First, many microorganisms form colonies or crusts on substrate components (sand grains and small stones). These colonies can be difficult to break apart, and fundamentally protect many cells from the lysis procedure. This can lead to differences in perceived community composition when comparing samples with different substrate grain sizes. Pre-washing procedures have been developed to suspend adhered communities in solution before breaking apart cells, leading to more representative extractions (Fortin et al., 2004; He et al., 2005).

Second, the thickness and material properties of cell walls and membranes are not uniform for all microorganisms. Developing lysis procedures therefore involves identifying the major groups of microorganisms of interest to the study, and considering performing parallel extractions to isolate distinct groups for comparison. Sulphate-reducing bacteria (SRB) and methanogenic archaea are of particular interest to the mining industry for their prevalence and activity in mined substrates. Bacterial and archaeal cell walls comprise different materials (see Kandler and König, 2014, for a chemical and functional description of these differences). As a result, using a single lysis procedure will invariably underrepresent one of these groups, and multiple lysis procedures should be evaluated if a comprehensive understanding of microbial community composition and prevalence is desired.

Finally, highly degraded organic material in carbon-rich soils (humic acids) are difficult to separate from DNA, often leading to contamination of DNA extracts and poor sequence reads (Yeates et al., 1998). Procedures have been developed to accommodate samples rich in organic acids; these involve removal of humic acids with additional filtration steps (Tsai and Olson, 1992) or ion exchange chromatography (Tebbe and Vahjen, 1993).
DNA extracted from the mine environment sample provides information about the microbial community and metabolic processes at that site. One method to describe the microbial community uses specific components of the DNA, such as genes known to vary among species, that are amplified via polymerase chain reaction (PCR) (Valentini et al., 2009). These reference genes found in the mine sample are compared to curated databases of known annotated sequences (BOLD, GenBank, MG-RAST) (Ratnasingham and Hebert, 2007; Meyer et al., 2008; Benson et al., 2013), thereby identifying which species are present at the mine site. Early uses of genomics technology for bacterial community characterisation focused on the 16S rRNA gene, a segment of DNA critical for the production of proteins by prokaryotic cells, and that is generally unique to each species (Weisburg et al., 1991). Sequencing of 16S rRNA genes has been used to quantify microbial diversity with sufficient resolution to detect shifts along major ecological gradients (Schmidt et al., 2014). However, this approach is limited to determining which microbial species are present and provides little information about the actual metabolic processes taking place (Eisen, 2007). Stable isotope (with 13C) probing has been used in recent years in combination with environmental DNA sequencing to isolate microbial community fractions involved in specific metabolic processes (Dumont and Murrell, 2005; Verastegui et al., 2014), which makes the important link between bacterial community composition and specific metabolic functions.

With the advent of high throughput sequencing technologies, targeting only specific regions within DNA has given way to “shotgun” and whole-genome/metagenome sequencing using all of the DNA in the sample (Tringe et al., 2005). The term metagenome is used when the sample contains many organisms (and thus genomes), as is the case in soils. Shotgun sequencing involves breaking the metagenome into many small fragments via physical shearing or enzymatic processes (Sharpton, 2014). After sequencing, these fragments can be used to reconstruct entire genomes for the organisms present at the site by aligning the sequences where they overlap. The metagenome includes both functional information (gene markers linked to specific metabolic products) and compositional information (e.g. 16S and other species-specific reference sequences) (Xia et al., 2011). The computational techniques required to compile, align and interpret these millions of bases of genetic code arranged into short fragments is termed bioinformatics (discussed in the next section).

An additional use of environmental DNA fragments is functional screening. Cloning of these fragments into live hosts such as the bacterium *E. coli* or the yeast *P. pastoris* enables expression of the protein products and their functional screening. For example, growing of *E. coli* cells containing environmental DNA on medium containing cellulose allowed for direct identification of novel cellulase enzymes (Mewis et al., 2013) from a mine remediation environment.

Bioinformatics refers to the computational procedures used to extract meaningful information from the very large datasets produced in metagenomic studies. Bioinformatics pipelines require considerable computational power and unique algorithms to carry out the steps from quality control of the sequences to assembling functional components (i.e. sequences coding for proteins used in cellular functions called open reading frames or ORFs) that yield information about the metabolic potential of the organisms present at the site. Bioinformatics for metagenomics are complex, as compositional as well as functional reconstruction is required. To identify compositional structure, the marker genes (i.e. 16S rRNA) are binned, based on their similarity, into operational taxonomic units (OTUs) that are likely to be derived from the same species. Assembled functional components (the ORFs) and OTUs are interrogated against protein and gene databases in bioinformatics pipelines using the blast tools (Altschul et al., 1990) in order to assign putative functions and identities to the microbial community found in the mine site sample (see Sharpton, 2014, for a comprehensive discussion of processing and interpretation of shotgun-derived sequences).
3 Applications

3.1 DNA barcoding

Given the much larger, more complex genomes found in plant and animal species, the molecular tools for identifying them have developed more slowly than those for microorganisms. Whereas the 16S rRNA gene is found in all bacteria and archaea, the corresponding 18S rRNA gene in higher eukarya, and even fungi, is not as reliable as a species identifier. As such, much effort is being put into finding signature genes, called “DNA barcodes,” for different groups of plants, invertebrates and vertebrates (Hebert et al., 2003). Each species collected can be rapidly and cost-effectively identified, and with the proper sampling protocols, quantified in terms of relative abundance and diversity by site.

New Gold’s New Afton Mine near Kamloops, British Columbia, Canada, began operation in 2012. Although the mine is in its early days of operation, it has taken a proactive approach to future mine closure plans by implementing a partnership with the Biodiversity Institute of Ontario to implement DNA barcoding for environmental impact assessments. The pilot program involved four sites: two grassland sites (disturbed and undisturbed) and two wetland sites (disturbed and undisturbed). Invertebrate samples from these four sites were collected in the summer of 2013, and DNA barcode analyses were completed in August 2014. Between 294 and 5,560 individual invertebrates were captured in Malaise traps each week, and 3,956 species were identified (D. Wilson and S. Davidson, personal communication 8 May, 2015). Differences were observed between habitat types such that wetlands contained more species than grasslands, and the natural grassland had more species than the disturbed grassland. The intention is to continue monitoring on a four- to five-year time scale. Such baseline data provide invaluable information for future mine closure and site reclamation.

3.2 Meta-omics

Thanks to the conserved nature of, and relatively long history of collecting sequence data for, the 16S rRNA gene in bacteria and archaea and, to a lesser extent, the 18S rRNA gene in fungi, researchers interested in microbial communities (e.g. environment, health, industry) have reached a stage where using these reference genes to describe microbial community composition and diversity is now routine (Schmidt et al., 2014). A short number of years ago, publications would be based solely on 16S rRNA gene surveys of microbial communities, whereas today these surveys are considered one of many standard analytical tools for scanning microbial landscapes. While 16S and 18S rRNA gene surveys will continue to be powerful tools, microbiologists have realised that these reference genes do not universally reflect the metabolic potential and biochemical activities of individual microorganisms, let alone complex microbial communities. This realisation continues to motivate the development of “meta” tools to qualify and quantify all of the DNA, mRNA and proteins in microbial communities (metagenomics, metatranscriptomics and metaproteomics) and to relate these data to biochemical fluxes (metabolomics) and, ultimately, ecosystem functions (Krause et al., 2014).

At the present time, there are good genomic databases for pure bacterial cultures grown under laboratory conditions. Generating good drafts of microbial genomes can now be done in days rather than years, by individuals rather than teams, for hundreds rather than millions of dollars. This is a major shift from less than a decade ago (Kyrpides et al., 2014). High throughput transcriptomic and proteomic tools have come online and are being increasingly used for pure culturable microorganisms as well. The tools are available to characterise natural microbial communities, but analysing data from even a few samples in a meaningful way still requires the use of supercomputers or powerful computer clusters, and is based on imperfect databases established in the early days of genomics (Howe et al., 2014). A “meta-omics” study is not for the faint of heart, but the field is shifting rapidly, and these types of studies are becoming more common despite the need for large monetary and personnel investments. Major breakthroughs are currently being made to advance our understanding of the dominant roles microorganisms play in the metabolism and lifestyles of all macroorganisms. Further, biotechnological applications are being realised in an array of fields, including
forestry, agriculture, animal husbandry, human health, and food, beverage and fuel production (Ekkers et al., 2012).

Given the rich literature on metal-microbe interactions (Gadd, 2010), there is tremendous scope for applying meta-omics to mine reclamation, particularly given the small but solid foundation of genomic work being done in phytoremediation (Bai et al., 2014), metal-plant interactions (Hanikenne and Nouet, 2011; Bhargava et al., 2012) and soil ecology (Howe et al., 2014) on which to build. Looking to the future, once current meta-omics tools have sufficiently matured for microbial communities, the gene, transcript, protein and metabolic signatures from plants and other macroorganisms must be integrated in holistic models in order to better appreciate the holobionts (macroorganisms and their associated microbial and viral communities) essential to healthy ecosystems.

3.3 State of the science

3.3.1 Example 1: Bioreactors

Seepage from mine tailings storage facilities, waste rock piles, open pits and underground workings, as well as excess process water, contains metals such as selenium, copper, molybdenum, zinc and arsenic, often along with sulphate and nitrate (McDonald and Strosher, 1998; Wang and Mulligan, 2006). High-density sludge treatment is used at many mine sites to treat this mine-influenced water, but this chemical process consumes large quantities of reagents and produces high volumes of toxic sludge requiring long-term safe storage (Zinck and Griffith, 2013). Bioreactors offer a sustainable, cost-effective alternative to reagent-based water treatment technologies.

Biological processes use natural microorganisms to treat mine-influenced water. Some bioreactors reduce sulphate to produce sulphide, which chemically binds with metal ions, causing them to precipitate out of solution as stable metal sulphides (Barnes et al., 1994). Additionally, bioreactors have been designed to selectively remove specific valuable metals, such as copper and zinc, that can be recycled to the metal extraction facility (Zinck and Griffith, 2013). Bioreactors are used to remove nitrate, which is high in some mine-influenced water owing to the use of explosives on mine sites (Koren et al., 2000). Selenium and arsenic are metals that occur as anions, and there are natural microorganisms that transform these compounds in order to gain energy for growth. Biological reactors using these organisms successfully remove selenium and arsenic from mine-influenced water (Morita et al., 2007).

Given the benefits in economics and effectiveness promised by biological treatment, it is surprising that use of bioreactors is not widespread on mine sites. A few reasons for this have been revealed through the application of metagenomics. Bioreactors contain consortia of microorganisms, rather than one single type (Baldwin et al., 2012), and the performance of these bioreactors depends on the types of microorganisms used to inoculate them (Pruden et al., 2007). Microbial communities are dynamic, and their members fluctuate in abundance and activity in response to changes in the influent water or operating conditions (Dar et al., 2008). Shifts in microbial community composition can lead to failure to meet water quality specifications, ultimately putting the receiving environment at risk (Mirjafari and Baldwin, 2011). To overcome this limitation of biological treatment, it is necessary to monitor microbial community composition in bioreactors and correlate this to operational settings and performance metrics. The microbial community in mine-based bioreactors can be monitored using the metagenomics techniques, such as those targeting the 16S rRNA, described in Section 2.3 (Schmidtova and Baldwin, 2011; Baldwin et al., 2015).

High-throughput sequencing can generate enough information to characterise the entire microbial community including very rare members. Often the important functional groups for metal removal in bioreactors are rare, even though their action achieves successful treatment (Rezadehbashi et al., 2012). Some other microbes present facilitate the activity of these desired groups, or some might compete for nutrients and hinder their activity. Nutrient consumption is one of the major costs in running a bioreactor, and when undesired microbes compete successfully for most of the nutrients, the desired organisms decline in number and bioreactor performance fails (Silva et al., 2012).
Much of the work done to date on microbial communities in bioreactors, especially those related to mine remediation, has focused on microbial community composition based on surveys of the 16S rRNA gene. These surveys have revealed that microbial groups associated with metal-rich environments are unknown and uncharacterised, such that new taxonomic groups have been invented to classify them (Khoshnoodi et al., 2013). Thus, metagenomic studies are needed to discover the potential functions of these novel organisms (Ellis et al., 2012).

Sequencing of DNA provides information on metabolic potential only, and does not reveal which of the genes are being actively expressed. Active metabolic processes can be detected by sequencing the transcribed genes using a technique referred to as transcriptomics (Luo et al., 2014). Although very recently developed for microbial communities in bioreactors (Luo et al., 2014), microbial transcriptomics is possible but requires careful sampling to preserve the active state of the microbes.

Given the affordability of high-throughput sequencing, use of metagenomics and transcriptomics for bioreactor monitoring is bound to increase. Since these data sets are new, correlation of microbial community composition, metagenomics and transcriptomics with bioreactor operation and performance is still being investigated. As the bioinformatics improves, these metagenomics tools show promise for future process control of bioreactors.

3.3.2 Example 2: Passive treatment systems

The benefit of using bioreactors for mine remediation is that they can be controlled using tanks, pumps, valves, instrumentation and defined nutrients to achieve just the right conditions and rapid kinetics for treating large flow rates (Zinck and Griffith, 2013). However, their capital and operating costs add to the expense of mine operation and closure. In addition, although they are highly automated, operators are needed.

The vast array of natural processes for biogeochemical cycling of metals and nutrients can be harnessed in so-called passive or semi-passive treatment systems (Ziemkiewicz et al., 2003). These typically take the form of constructed wetlands, either sub-surface flow anaerobic, surface flow aerobic or, most commonly, combinations thereof. Instead of defined nutrients, mixtures of waste organic materials such as wood debris, hay, compost, pulp and paper mill biosolids are used. These complex organic materials are decomposed into the smaller carbon compounds needed for the sulphate- and metal-reducing bacteria.

If successful, passive treatment can remove metals at seeps distributed across the mine site for a fraction of the cost of active bioreactor treatment (Zinck and Griffith, 2013, p. 14). The metal residuals are captured and secured inside the organic matrix, most often as sparingly soluble metal sulphides (Khoshnoodi et al., 2013). The ecosystem of a passive remediation system is as complicated as that of soils (Baldwin et al., 2015). Like bioreactors, they are consortia of interacting species that shift in composition with geochemical gradients, seasons and the age of the system (van der Lelie et al., 2012). Their performance may decline as the organic material decomposes, evolving towards microbial communities with completely different metabolic potential than at the start (Mirjafari, 2014). Preliminary studies of the microbial communities in these systems reveal that the metabolic potential for metal removal in them is much wider than previously thought (Baldwin et al., 2015). They contain species that tolerate high metal concentrations, many novel unclassified candidate division groups found in other metal-contaminated environments and species capable of using usually recalcitrant aromatic compounds for growth.

Successful metal removal occurs even in places where SRB are extremely rare, meaning either that many other groups of organisms we do not know about are capable of sulphate reduction and/or metal precipitation, or that only a few sulphate reducers are needed for successful treatment (Khoshnoodi et al., 2013; Baldwin et al., 2015). Charting of microbial communities in passive remediation bioreactors has revealed that they are not static, but fluctuate cyclically (Baldwin, unpublished data). Using metagenomics and metatranscriptomics, we can learn more about the dynamics of these ecosystems as they respond to changing conditions and use this knowledge to design better systems or diagnose performance issues.
3.3.3 Example 3: Novel gene discovery

Metal-rich ecosystems are considered extreme environments. They harbour highly specialised microorganisms that have evolved unique metabolisms to transform, sequester or detoxify metals in order to survive. Examples include sulphate reducers’ overproduction of extracellular polymeric material to bind up copper ions, thereby creating nucleation sites for precipitation (Jalali and Baldwin, 2000), intracellular mineralisation of tellurium to sequester this highly toxic metal (Amoozegar and Khoshnoodi, 2012), and methylation and volatilisation of arsenic by *Methanocorpusculum labreanum*, suggesting that it may be a significant contributor to metal cycling in anaerobic environments (Khoshnoodi et al., 2012).

Enzymes involved in metal cycling, or biochemical compounds with the ability to sequester specific metal ions, can be used in future biotechnologies to improve bioremediation, and even develop methods for in situ mining. The field of functional metagenomics, first mentioned in Section 2.3, is being used to screen large DNA fragments from mine sites (Mewis et al., 2011). These large fragments of environmental DNA may contain novel genes for metal cycling, and, using selective media in the laboratory, we can screen the *E. coli* clones for metal resistance. It may be possible in the future to construct biochemical pathways for metal removal using simple and easy-to-grow organisms and synthetic biology.

3.4 Future of the science

3.4.1 Site-specific pre-assessments for closure targets

The mining industry has recognised for some time that planning for mine closure begins even before overburden is removed from the site (Thirgood, 1986 in Polster 1989). Until recently, this has not included a significant consideration of the pre-impact community of plants, animals and microorganisms residing in the unaltered substrate, but see Morrison et al. (2005) and Jasper (2007). In the future, such assessments should include evaluation of both unimpacted overburden soil communities and stockpiles of such material for eventual recovering of the site. Microbial genomics may play a key role in expediting mine land restoration by providing information on soil community dynamics in overburden stockpiles, allowing managers to maintain these communities in a way that expedites recovery once these substrates are used to recover the mined site.

Microbial genomics may also serve as an indicator for success in these systems, providing additional evidence of soil community formation and ecosystem trajectory. To be able to set objectives for effective mine reclamation and evaluate what is successful, the meaning of “equivalent land capacity” needs to be defined for every mine site.

Barcoding and metagenomics can be incorporated into methods for evaluating ecosystem services. Microbial community analysis provides valuable information about nutrient (carbon, nitrogen and sulphur) cycling, greenhouse gas emissions and metal transformations that can be fed into determining pre-mining land capacity assessment. This will allow targets to be specified for post-mine reclamation, and these same tools can be used to evaluate whether remediation strategies are working.

3.4.2 Whole-ecosystem modelling

Microbial genomics fills a critical gap in efforts to simulate the formation of metabolic networks in ecosystems. The presence or absence of key microbial communities may mean the difference between a successfully remediated site and persistent degradation of water quality (e.g. acidification and metal leaching). Efforts are underway to develop comprehensive predictive networks for mine sites, in which environmental genomics are tied to environmental monitoring data to generate a comprehensive understanding of ecosystem function. Probabilistic modelling frameworks may very well provide early warning signs of acid generation in mined substrates, and genomics data may yield the necessary evidence to determine what remediation options will be most likely to succeed in both the short and long term.
3.4.4 Defining benchmarks

Defining benchmarks may very well be the key component necessary to leverage genomics tools for mine-site management. Bioinformatics data processing workflows (pipelines) exist (sometimes incorporated directly into sequencing hardware) that provide for rapid processing and interpretation of sequencing runs. The great questions for the mining industry can be put quite simply: (1) What do we sequence?, and (2) How do we use the resulting data to improve management, reclamation and containment?

Section 2.1 provides guidelines for sampling of these sites, but the industry requires a wealth of context-dependent validation of these approaches to become widespread. Therefore, research and development resources (both academic and industrial) should be leveraged in such a way that sampling procedures can be standardised to the sorts of questions of interest at mine sites (e.g. How do soil microbial communities form on mined substrate? What soil conditions promote recovery of native species?).

Ultimately, with standard sampling procedures, bioinformatics pipelines can be tailored to industry-specific questions. For example, a sampling protocol might be developed to evaluate diffuse leachate exposure in riparian substrates, with the question: Does exposure to leachate affect microbial community composition and dynamics in a way that could limit containment in the future? Field sampling and sequencing following established protocols would be processed through a tailored bioinformatics pipeline that yields the following outputs:

1. Community composition matrices and planned comparisons (diversity, similarity, etc.)
2. Compositional differences and known associations (i.e. variations in groups of organisms known to be involved in certain geochemical pathways)
3. Management recommendations and references to similar scenarios/responses.

4 Conclusions

We have outlined a range of genomics applications to mine closure, from characterising natural substrates before overburden removal, to water treatment, to bioremediation and monitoring of healthy reclaimed ecosystems. In performing this review, we identified two key constraints on the widespread application of genomics for mine closure. First, industry-wide standard operating protocols need to be developed for mine closure, including sampling procedures designed for representativeness and comparability (spatial extent, replication, temporal frequency). Second, sequencing and data interpretation pipelines must be established in parallel with the development of these standards, allowing mine managers to more easily discover what works in a given system and to establish benchmarks for reclamation success. Mining is a key global industry for development and for quality of life. The integration of genomics technologies into mine closure planning and implementation may drastically improve the stability and reliability of ecosystem reclamation.

Acknowledgements

We would like to thank Jon Taylor (University of British Columbia) for helpful troubleshooting and advisement on sampling design for mine-site applications. We would like to thank Genome BC, Genome Canada, NSERC Engage and Mitacs Accelerate for funding to support preliminary R&D on metagenomics technologies. We would also like to thank Dennis Wilson and Scott Davidson of New Gold for providing details about their project. Finally, we would like to thank Mount Polley Mine Corporation and particularly Katie McMahon, Colleen Hughes, Luke Moger and Art Frye for their partnership in projects aimed at addressing many of the questions addressed in this review.
References


