Next-Generation Sequencing of Protists as a Measure of Natural Soil Microbial Eukaryotic Community in the Oil Sands Region

Elisabeth Richardson[1*], G. Walker[1*], Georgina MacIntyre[2], S. Quideau[3], J. B. Dacks[1] and S. Adl[4]

[1] Department of Cell Biology, University of Alberta

[2] The Applied Genomics Core (TAGC), University of Alberta

[3] Department of Renewable Resources, University of Alberta

[4] Department of Soil Sciences, University of Saskatchewan

* denotes equal contribution

December 2014

Oil Sands Research and Information Network

The Oil Sands Research and Information Network (OSRIN) is a university-based, independent organization that compiles, interprets and analyses available knowledge about managing the environmental impacts to landscapes and water impacted by oil sands mining and gets that knowledge into the hands of those who can use it to drive breakthrough improvements in regulations and practices. OSRIN is a project of the University of Alberta's School of Energy and the Environment (SEE). OSRIN was launched with a start-up grant of $4.5 million from Alberta Environment and a $250,000 grant from the Canada School of Energy and Environment Ltd.

OSRIN provides:

- **Governments** with the independent, objective, and credible information and analysis required to put appropriate regulatory and policy frameworks in place
- **Media, opinion leaders and the general public** with the facts about oil sands development, its environmental and social impacts, and landscape/water reclamation activities – so that public dialogue and policy is informed by solid evidence
- **Industry** with ready access to an integrated view of research that will help them make and execute environmental management plans – a view that crosses disciplines and organizational boundaries

OSRIN recognizes that much research has been done in these areas by a variety of players over 40 years of oil sands development. OSRIN synthesizes this collective knowledge and presents it in a form that allows others to use it to solve pressing problems.

Citation

This report may be cited as:

Richardson, E., G. Walker, G. MacIntyre, S. Quideau, J.B. Dacks and S. Adl, 2014. Next-Generation Sequencing of Protists as a Measure of Natural Soil Microbial Eukaryotic Community in the Oil Sands Region. Oil Sands Research and Information Network, University of Alberta, School of Energy and the Environment, Edmonton, Alberta. OSRIN Report No. TR-69. 26 pp.

Copies of this report may be obtained from OSRIN at osrin@ualberta.ca or through the OSRIN website at http://www.osrin.ualberta.ca/en/OSRINPublications.aspx or directly from the University of Alberta's Education & Research Archive at http://hdl.handle.net/10402/era.17507.

Table of Contents

LIST OF TABLES iii

LIST OF FIGURES iii

REPORT SUMMARY iv

ACKNOWLEDGEMENTS v

1 INTRODUCTION 1

1.1 Importance of Soil Biodiversity 1

1.2 Theoretical Benefits of using the Illumina MiSeq Reagent Kit V2 (500-cycles) versus Illumina MiSeq Reagent Kit V3 (600-cycles) 2

1.3 Research Objectives 3

2 EXPERIMENTAL APPROACH 3

2.1 Site Selection and Soil Sampling 3

2.2 DNA extraction and sequencing 4

2.3 Preliminary Analyses of DNA Sequences 5

3 RESULTS 5

3.1 Comparison of 2 x 250 bp vs 2 x 300 bp kits for a Single D1 Library 5

3.2 Replicates of D1 8

4 DISCUSSION 13

4.1 V2 2 x 250 bp vs V3 2 x 300 bp Kits 13

4.2 Biodiversity and Taxa Present 13

5 CONCLUSIONS 14

6 REFERENCES 15

7 GLOSSARY 17

7.1 Terms 17

7.2 Acronyms 19

LIST OF OSRIN REPORTS 20

LIST OF TABLES

Table 1. Run quantities from D1 replicate samples. 8

LIST OF FIGURES

Figure 1. Diagram of amplicon and paired end strategies. 3
Figure 2. Relative performance of data quality measures. 7
Figure 3. Biodiversity of D1 samples. 10
Figure 4. Biodiversity of the samples, with Metazoa, Embryophyta and Fungi removed. .. 11
Figure 5. Biodiversity of the samples (fungi only). 12

REPORT SUMMARY

Soil plays a central role in the functioning of all terrestrial ecosystems. Among the many ecosystem services to which soil contributes are: purification and storage of water, sequestration of organic matter, nutrient cycling for plant growth, and conservation of healthy faunal and microbial populations. As such, soil quality is a key determinant of land reclamation success. Exploitation of the Athabasca oil sands deposit represents a massive land disturbance in Alberta. To date, over 600 km^2 of land has been disturbed by oil sands extraction. Following surface mining, reclamation efforts involve the reconstruction of entire landforms. Salvaged surface soils and near-surface geological materials are placed as a new soil cover on the reconstructed landscapes. The goal of reclamation in Alberta is to achieve land capability equivalent to that which existed prior to disturbance. Soil parameters that are currently used to examine reclamation success include chemical and physical attributes known to limit plant growth. Although it is essential to the functioning of these reconstructed ecosystems, soil biology is not yet included as part of the assessment.

This project characterized for the first time the biodiversity of soil microfaunal and mesofaunal populations on natural Athabasca oil sands sites. Specifically, we focused on soil protists and micro-invertebrates, as these largely bacteria-consuming organisms are responsible for much of the nutrient fluxes through the soil food web and have crucial bottom-up impact on animal and plant biodiversity.

The report addressed two issues. The first is to pilot the use of Next-Generation Sequencing (NGS) technology to establish an assessment of soil protist and invertebrate biodiversity in undisturbed soils as a starting point to identify bioindicators for future assessments of reclamation success. The second question was to assess the relative merits of using paired end 250 versus paired end 300 kits in the NGS protocol, as a technical note going forward.

We found that, for these samples, the paired end 250 kits and the protocols in place to use them were reliable and produced consistent datasets and were sufficient to capture the diversity within our samples. Therefore, the additional cost of the paired end 300 kit was not warranted for our needs and would not be adopted in future NGS studies for these organisms in this particular environment.

This first assessment of soil protist biodiversity revealed similar trends to those seen from other NGS studies of soils, with cercozoans and ciliates as obvious components of the biodiversity. Further quantitative analysis is key to making any statements about the numerical abundance of any taxa in our samples. However, there are no obviously comparable samples available, with the closest analysis of boreal forest soil being performed with key differences in NGS technology, and the closest technically comparable sample coming from conifer soils in the southern USA. Although we have made as relevant comparisons as possible, to our knowledge this represents the first report of microbial eukaryotic biodiversity of undisturbed soil in the Athabasca region and is an important first step in assessing downstream efforts for soil reclamation and revegetation.

ACKNOWLEDGEMENTS

The Oil Sands Research and Information Network (OSRIN), School of Energy and the Environment (SEE), University of Alberta provided funding for this project. We also gratefully acknowledge practical help from Mathew Swallow (University of Calgary), and analytical help from Marieta Aguilar Gonzalez (University of Alberta).

1 INTRODUCTION

Soil plays a central role in the functioning of all terrestrial ecosystems. Among the many ecosystem services to which soil contributes are purification and storage of water, sequestration of organic matter, recycling and release of nutrients for plant growth, and conservation of healthy faunal and microbial populations (Bardgett 2005). As such, soil quality is a key determinant of land reclamation success, and the microbiota living in soil determine its quality, permitting microbiological approaches to reclamation (Adl 2008, Dimitriu and Grayston 2010). Exploitation of the Athabasca oil sands deposit represents a tremendous land disturbance in Alberta. To date, over 600 km^2 of land have been disturbed by oil sands extraction. Following surface mining, reclamation efforts involve the reconstruction of entire landforms. Salvaged surface soils and near-surface geological materials are placed as a new soil cover on the reconstructed landscapes. The goal of reclamation in Alberta is to achieve land capability equivalent to that which existed prior to disturbance. In a report commissioned by the Reclamation Working Group of the Cumulative Environmental Management Association (CEMA), proposed indicators for evaluating the success of reclamation include ecosystem net primary productivity; plant community composition, diversity, richness and abundance; and soil fungal mycorrhizal communities (Poscente and Charette 2012). While these indicators are not yet used, and mycorrhizal community characteristics require further research before being used, they depend directly on the health of soil microbiological communities (Adl 2008, Poscente and Charette 2012).

1.1 Importance of Soil Biodiversity

Soil biota, directly or indirectly, drive all of the goods and services provided by terrestrial ecosystems. Soil biodiversity is increasingly acknowledged as one of the key attributes of soil health, and its decline has been identified as one of the eight major current threats to soil by the European Commission – Joint Research Centre Institute for Environment and Sustainability, together with the loss of organic carbon in topsoils, and salinization (Adl 2008). In Alberta, potential soil issues arising during reclamation including decreases in nutrient stocks or salinity are clearly identified in site assessment guidelines (Alberta Environment 2006). Yet, the biological mediation of these issues is currently not included in required assessments, and is at a relatively early stage of investigation (Dimitriu and Grayston 2010), with protists or micro-invertebrates not yet represented in the literature.

Significant work has been done on the biodiversity of microbial eukaryotes in soil using microscopy and culturing methods (e.g., Adl and Gupta 2006). They inhabit multiple niches in the soil ecosystem, largely as bacteriovores (e.g., cercozoa, and many amoeboid species), or predators at other trophic levels such as cytotrophs (ciliates, some testate amoebae) or fungivores (some ciliates and many amoeboid species) but also as primary saprotrophs (most fungi and some oomycetes). Soil is also a reservoir for spores of parasites of plants and animals, as well as symbionts of plants (mycorrhizae). With between 10^5 and10^8 protist cells active in a given gram of soil and debris, the role that protists play is significant, as "protists in the food web help return CO_2 to the air for photosynthesis, return soluble nutrient molecules to the soil solution for root

uptake, and excrete undigested food vacuole contents as humus for further digestion in the soil food web" (Adl and Gupta 2006). As for invertebrates in the oil sands region, this is the first comprehensive assessment of their diversity.

An alternate approach to culturing and microscopy is to use Next-Generation Sequencing (NGS) technology to do molecular surveys of diversity and community structure within particular marine (e.g., Lovejoy 2006), freshwater (e.g., Richards 2005) and even tailings pond environments (e.g., Aguilar et al. 2014, Aguilar et al. submitted). Recent analyses using NGS technology has revealed cryptic diversity in soil protists and provided more detailed enumeration of community composition and diversity (e.g., Baily 2007). Several key papers in the last two years did comparative studies of microbial eukaryotes generally (Bates et al. 2013) or of targeted groups, such as Fungi (Tedersoo et al. 2014) in comparable environments to those we are assessing. These will serve as important comparison points for this study.

1.2 Theoretical Benefits of using the Illumina MiSeq Reagent Kit V2 (500-cycles) versus Illumina MiSeq Reagent Kit V3 (600-cycles)

A common strategy for estimating microbial biodiversity using NGS is an "amplicon-based approach". The polymerase chain reaction (PCR) is used to amplify a specific region of a gene of interest. The most commonly utilised gene is the small subunit ribosomal DNA gene (SSU rDNA), due to its historical use as a taxonomic marker, resulting in the most comprehensive databases against which to compare sequences from new environments. Because PCR relies on primers to amplify the chosen region, the DNA sequence of the primers can be tailored to the taxonomic scope of interest, whether to a specific group (bacterial, fungal) or, in our case, all eukaryotes. The amplified regions of the gene (i.e., amplicons) are then sequenced using the NGS technologies.

One of the methodological limitations to this approach is that the amplicons may be longer than the reads from the sequencing reaction. This is exacerbated because the same region of the SSU rDNA gene may be variable in length between different organisms in the community that is being examined. As the technologies are improving to produce longer sequencing reads of DNA this problem is becoming less of a concern, but it is still important to take its effect into account. This is somewhat mitigated by the use of a "paired-end" strategy. In this protocol, unique sets of nucleotides of known sequence are placed on the ends of the amplicons. This allows for reads going from each direction to be matched up based on the sequence of the DNA in the regions that overlap (see Figure 1). Nonetheless, there are two immediate technical issues to consider. Firstly, if a given amplicon is too long, then the paired end reads may not overlap sufficiently to allow for enough information to match the correct sequences together into 'contigs". Secondly, there is always an error rate associated with the matching process resulting in two sequences being assembled in to a "chimaera" that do not represent a true gene.

The fragment of the SSU rDNA gene that we chose to amplify, the V4 region, is at a critical point in this consideration as it is typically 400 bp, but can range beyond that in some known eukaryotes. The Illumina MiSeq platform on which we have chosen to perform our NGS data collection offers two options for paired end sequencing for this DNA length; MiSeq Reagent Kit

V2 (500-cycles) (henceforth abbreviated as 2 x 250 bp) that can sequence read 2 x 250 bp in length and MiSeq Reagent Kit V3 (600-cycles) (henceforth abbreviated as 2 x 300 bp) that can sequence read 2 x 300 bp in length. Additionally, due to the new V3 chemistry of the 2 x 300 kits, there is a theoretical increase in the overall data collected (up to 15GB and 44 to 50M reads for 2 x 300 bp compared to 8.5 GB and 24 to 30M reads for 2 x 250 bp with the V2 chemistry). Theoretically we may observe an improvement using the 2 x 300 bp kit over the 2 x 250 bp kit due to the increased length of the sequencing reads, improved quality scores and the overall amount of data collected.

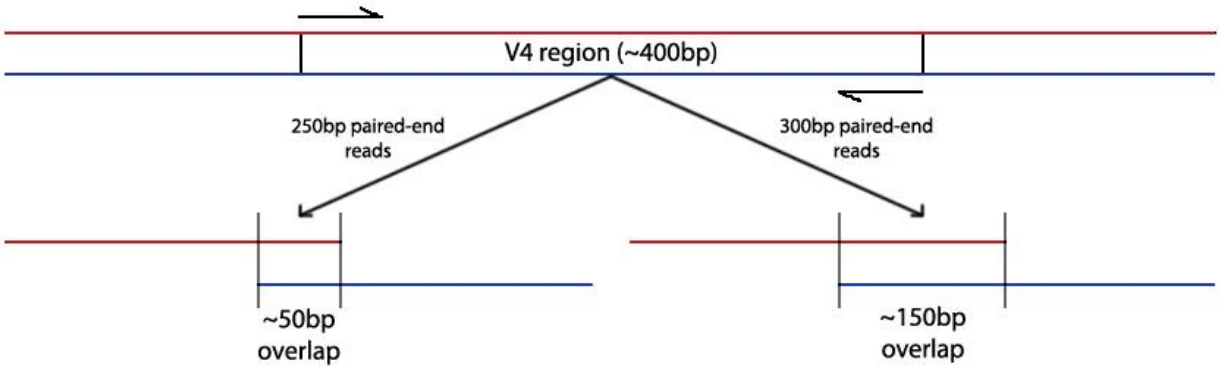

Figure 1. Diagram of amplicon and paired end strategies.

1.3 Research Objectives

We know from research at other reclaimed sites that a central issue for reclamation is establishment of the soil food web, and site preparation is key (Frouz 2013). Our end-goal is to identify protists, fungi and invertebrate animals that are early indicators of reclamation success. In the short-term our goal is to:

1. obtain a first assessment of naturally occurring microbial eukaryotic diversity in undisturbed soil from the oil sands region, and
2. assess the best protocol and reagents to use in NGS exploration of microbial diversity in this environment.

2 EXPERIMENTAL APPROACH

2.1 Site Selection and Soil Sampling

We took advantage of a network of sites that we had been previously examined (Sorenson et al. 2011, Turcotte et al. 2009) to select a natural site (D1). This site, which has the characteristics of a d1 ecotype as described in Beckingham and Archibald (1996), was sampled in September 2013 and aliquots were kept refrigerated (but not frozen) until analysis. The site is found at latitude

57° 08’ 51.4’’ and longitude -111° 32’ 38.40’’. At this experimental site, four locations were chosen at the four cardinal directions and a ten-metre distance from a central point. At each location, the organic layer and the top 5 cm of mineral soil were sampled separately. Only the organic samples (D1-org) are discussed in this report. About 100 mL of soil was collected for each replicate in a sterile Ziploc plastic bag, attempting not to disturb the internal structure of the sample.

2.2 DNA extraction and sequencing

Four DNA extractions were performed per sample on 0.5 g of unsieved, unfiltered soil, using the MoBio Powersoil Kit following the manufacturer’s instructions (MoBio Laboratories Inc., Carlsbad, California, USA). For each DNA extraction, four replicate PCR amplifications of the V4 region of the SSU rRNA gene were performed, using general eukaryotic primers TAReuk454FWD1[1] and TAReukREV3[2] (Stoeck et al. 2010) and the Thermo Scientific Phusion Hot Start II High Fidelity DNA polymerase kit (Fisher Thermo Scientific, Pennsylvania, USA). The primers also contained overhang linker sequences for Illumina MiSeq two-step amplicon library building, using manufacturer’s instructions for the overhang sequences[3]. Although we sampled N,S,E,W from each sampling site, for sequencing only three were used, according to the PCRs that gave the most product of the correct size. These correspond to the D1-org, D1-O2 and D1-O3 libraries.

PCR products were pooled from all four PCR reactions, and purified using the Agencourt AMPure XP system (Beckman Coulter, California, USA). Cleaned products were quantified on a Bioanalyzer 2100 High Sensitivity DNA chip using the DNA 1000 kit (Agilent Technologies, California, USA). Libraries were prepared for each sample using unique barcode adapters, using the Illumina Nextera XT kit (Illumina, California, USA), assayed further to determine the final concentration for sequencing using the Bioanalyzer 2100 chip, and sequenced as 1/12th of a single Illumina MiSeq run (250 bp paired end reads) by TAGC[4], University of Alberta. For comparative purposes, an aliquot of the same library was sequenced as 1/12th of a 300 bp paired end read run. Some recommendations regarding sample quality control and run parameters have also changed since the initial V2 2 x 250 bp runs; these include the use of the Qubit fluorometer (Life Technologies), rather than the BioAnalyzer (Agilent), for quantification of DNA libraries, and changes to the amount of PhiX loading control applied per run (20% for previous runs and 1% to 5% currently). As previously, the BioAnalyzer (Agilent) was used to check the quality of the libraries and to determine the average length of DNA fragments within each library.

---

[1] 5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCC-AGCASCYGCGGTAATTCC-3'

[2] 5'-GTCTCGTGGGCTCGGAGATG-TGTATAAGAGACAGACTTTCGTTCTTGATYRA-3'

[3] See http://res.illumina.com/documents/products%5Cappnotes%5Cappnote_16s_sequencing.pdf

[4] See http://tagc.med.ualberta.ca/

2.3 Preliminary Analyses of DNA Sequences

Data were returned as a FASTQ file of forward or reverse read sequences for each of the two runs. Further sequence analysis was carried out using the microbial ecology software package mothur, following the MiSeq analysis pipeline outlined at http://www.mothur.org/wiki/MiSeq_SOP.

Contigs were assembled from paired-end reads into a single FASTA file of complete sequences. Sequences were filtered to remove ambiguous reads and any longer than 600 bp. Non-unique sequences were counted and eliminated; this step resulted in the formation of a table of unique sequence identifiers and the number of that particular sequence within the original file.

Unique sequences were aligned against a reference for the same 18S V4 amplicon region; the reference used was the Silva eukaryotic ribosomal database (obtained from http://www.arb-silva.de/). Sequences that did not align to the correct region were discarded. Sequences with a large number of homopolymers (defined for this analysis as more than 8) were also discarded. Sequences were aligned against each other to further verify the correct region was present, and, using the most abundant sequence as a reference, chimeric sequences were removed by the uchime package.

Once the final filtered sequence file was obtained, sequences were classified using the Silva eukaryotic reference taxonomy. Any sequences aligning to archaea, bacteria, chloroplasts or mitochondria were discarded, as well as unknown sequences.

More detailed classification of the resulting FASTA files was obtained using the PR2 ribosomal reference database (Guillou et al. 2013), which contains sequence data at much higher taxonomic resolution than the Silva database used in mothur. Each sequence was BLASTed against the PR2 database for classification, and information was collected to the phylum level. The BLAST algorithm[5] also provides information on how similar the query sequence is to the reference sequence, and this was used to discard any poorly classified sequences (defined as those with an E value greater than 0.05).

After classification, the number of unique sequences in each phylum was counted to determine the distribution and abundance of phyla. The diversity of the sample was also assessed using the standard diversity indexes of species richness, Shannon diversity, and Simpson diversity. These were calculated using the vegan package in R.

3 RESULTS

3.1 Comparison of 2 x 250 bp vs 2 x 300 bp kits for a Single D1 Library

The choice of technical protocol for sequencing can vary between environments and depend on the diversity of eukaryotes and the presence of particular organisms, and thus the characteristics of their genes of interest. To assess the relative merits of using the 2 x 250 bp or the 2 x 300 bp

[5] See http://blast.ncbi.nlm.nih.gov/Blast.cgi

kit on our samples from this environment, we sequenced the same library (sample name D1-org) derived from an undisturbed boreal forest organic layer from the oil sands region using both kits.

This resulted in 1,148,793 reads using the 2 x 250 bp kit and 950,623 reads using the 2 x 300 bp kit. The reads from each run were processed and analyzed bioinformatically as described above. Of specific interest we assembled the reads into contigs, and then performed several quality control steps. This involved removing ambiguous reads, reads that could not be confidently aligned to the reference database (in this case the SILVA database), reads that were likely due to sequencing errors, and chimaeras. The total number of reads was also counted separately from the number of unique reads. Using these steps, we then assessed the number of sequences to test three theoretical advantages of the 2 x 300 bp kit over the 2 x 250 bp kit.

Given that the V4 region of the SSU rDNA gene can vary up beyond 400 bp in length, if there are amplicons that are so long as to be beyond the limit of paired end reads of 250 bases, this should result in some of the non-overlapping reads (they may still be paired – but no overlap) in a sequencing run using the 2 x 250 bp kit that would be more confidently assembled by the 2 x 300 bp kit. However, in both of the sequencing runs, the pairing of reads was 100%. Therefore, the 2 x 250 bp kit appears to perform adequately for our samples in this instance.

For these amplicons, the 2 x 300 bp kit should allow us to identify more of the organisms with a V4 region longer than 500 bp, whereas they may be missed using the 2 x 250 bp kit. Thus the 2 x 300 bp kit should reveal some novel sequences as compared to the 2 x 250 bp kit. This was observed, but in a run of 76,918 classified reads only nine new sequences were found. These sequences, when classified, were additional sequences of classes already seen in the 2 x 250 bp run. Therefore the overall assessment of community structure does not change upon the inclusion of these sequences.

For amplicons where overlap does exist using both kits, the incorrectly matched reads should be reduced in the 2 x 300 bp kit, due to the ~100 bp of additional overlap. This should theoretically translate to a reduction in chimaeras in the 2 x 300 bp vs 2 x 250 bp run. As seen in Figure 2 this was not observed with 40% of the total sequences removed in the 2 x 250 bp and 46% in the 2 x 300 bp.

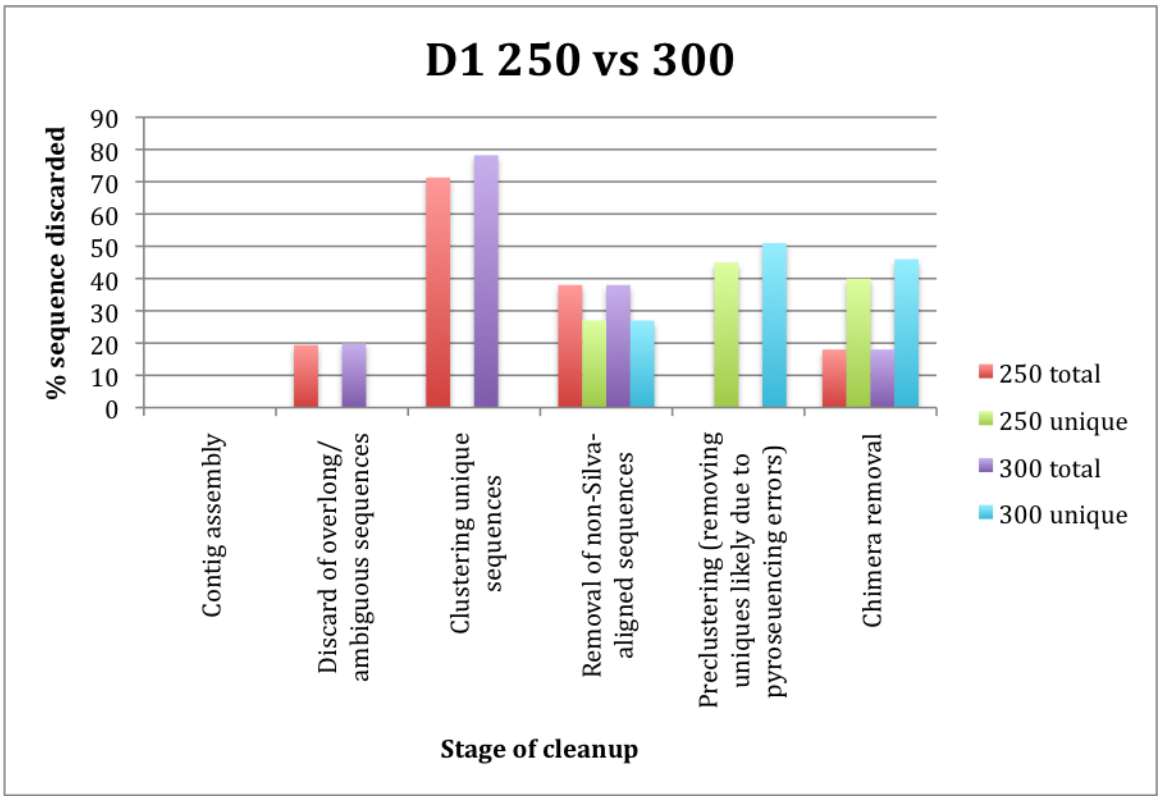

Figure 2. Relative performance of data quality measures.

One of the stated advantages of the 2 x 300 bp kit is the increased amount of sequence collected. Therefore our final assessment was the total amount of sequence that was obtained in the two runs for this library. We were surprised to see that the 2 x 300 bp run involved fewer sequences than the 2 x 250 bp. Since both libraries were calibrated to be 1/12th of the total run, we therefore assessed the total number of reads and data collected by the MiSeq runs on the dates that the data were collected (Table 1). We noted that the runs of the libraries made from the physical sampling replicates of the D1 environment were within 2% of one another when sequenced using the 2 x 250 bp kit, but ranged from 950,623 to 2,653,735 (more than 250%) when using the 2 x 300 bp kit. This was despite the fact that the 2 x 250 bp runs were performed on different dates and with different libraries in the MiSeq runs, while the 2 x 300 bp libraries were run together on the same run. The wider range of reads generated is likely due to the use of the Qubit to measure DNA concentrations. The DNA concentration estimates for each of the three samples was higher using the Qubit than found previously using the BioAnalyzer (e.g., D1-org 53.59 nM by BioAnalyzer and 163.3 nM by Qubit). This resulted in a three-fold higher estimate for concentration for D1-org and a two-fold higher estimate for D1-O2 and D1-O3.

Table 1. Run quantities from D1 replicate samples.

		D1-org	D1-O2	D1-O3
250.00	Reads	1,148,793	1,145,282	1,167,174
	Mean length	377.89	339.83	337.72
	Median length	418	417	418
	Mbp	434.11	389.20	394.17
300.00	Reads	950,623	2,360,551	2,653,735
	Mean length	375.51	360.02	357.57
	Median length	418	418	418
	Mbp	356.97	849.84	948.88
	Reads ratio	1.21	0.49	0.44
	Mbp ratio	1.22	0.46	0.42

Overall, our assessment was that in the case of the D1-org sample, the change in recommendations related to quality control resulted in an over-estimate for DNA concentration leading to a lower amount of DNA applied to the MiSeq run, with a subsequent reduction in the total number of reads obtained for this sample compared to the previous V2 2 x 250 bp run. For D1-O2 and D1-O3 the number of reads obtained was more than doubled using the V3 2 x 300 bp kit, as expected. For all samples, mean lengths were only slightly improved and median lengths remained the same. The changes to the PhiX concentration were not detrimental. The replicate data from the V2 2 x 250 bp was the most consistent and these data were used for the comparison of the community structure.

3.2 Replicates of D1

To assess the community structure and biodiversity of the D1 soil sample, the contigs from the 250 runs of the D1-org and D1-O2 replicates were classified using BLAST against the SILVA database. We note that due to time limitations and computational tractability, we did not perform OTU clustering, nor were we able to do final classifications for the D1-O3 sample and so the abundance is a measure of sequences at a given taxonomic classification, and not abundance at the OTU level. As well, the totals could be influenced by the presence of uneven sampling (e.g., fragment of a multicellular organism). We wish to clarify that the relative numbers here are on sequence counts only, and will need to be reassessed once OTUs are completed. However, they do give some assessment of the data obtained and are useful, when the caveats are taken into consideration.

Overall, we found our samples to be dominated by metazoa (annelida, gastrotricha and nematoda), and fungal (Mucoromycota) sequences (Figure 3). This could be due to uneven sampling of a multicellular organism. Nonetheless, microbial eukaryotes were amongst the five most abundant taxonomic groups seen with cercozoans[6] being between 5.9% and 10.2%. We calculated the Shannon and Simpson indices of 2.8 and 0.88, respectively, indicating unevenness in our community structure, although again this would need to be reassessed upon obtaining OTU counts. While there were taxa that clearly dominated the sequence totals, the samples were also highly diverse with 106 to 111 unique sequences in our samples. The rare biosphere can be important reservoirs of ecological potential in a given environment, as these may become dominant in different ecological conditions (temperature or moisture content). Indeed we found between 58 and 69 "rare" sequences in our samples as defined by being represented more than once, but less than 0.1% of the total.

As our analysis was aimed at understanding microbial eukaryotic diversity, we assessed the relative diversity of the sample having discounted the sequences assigned to metazoa and embryophyte plants[7], as well as Fungi.

As seen in Figure 4, our samples were dominated by ciliate and cercozoans sequences, common and abundant taxa found in soils globally. Ochrophyta[8] were also found abundantly in both samples. Additional abundant taxa were amoebozoa[9] (Conosa and Lobosa), as well as apicomplexa[10]. The latter are dispersal spore stages as apicomplexans are exclusively parasites of Animalia and a few protist lineages. The Shannon and Simpson indices were ~2.75 and 0.86, respectively. There was a clear tail of rare sequences, with between 28 and 43 taxa accounting for ~1% of the total sequence numbers.

Although the focus of our initial study was microbial eukaryotes, Fungi are known to be an important component of the soil microbiota. Therefore we also performed a separate analysis of the fungal diversity in our samples. We note however, that fungal diversity studies typically use SSU-ITS regions.

---

[6] See http://tolweb.org/Cercozoa/121187

[7] Terrestrial plants

[8] See http://www.inaturalist.org/taxa/48221-Ochrophyta

[9] See http://en.wikipedia.org/wiki/Amoebozoa

[10] See http://www.ucmp.berkeley.edu/protista/apicomplexa.html

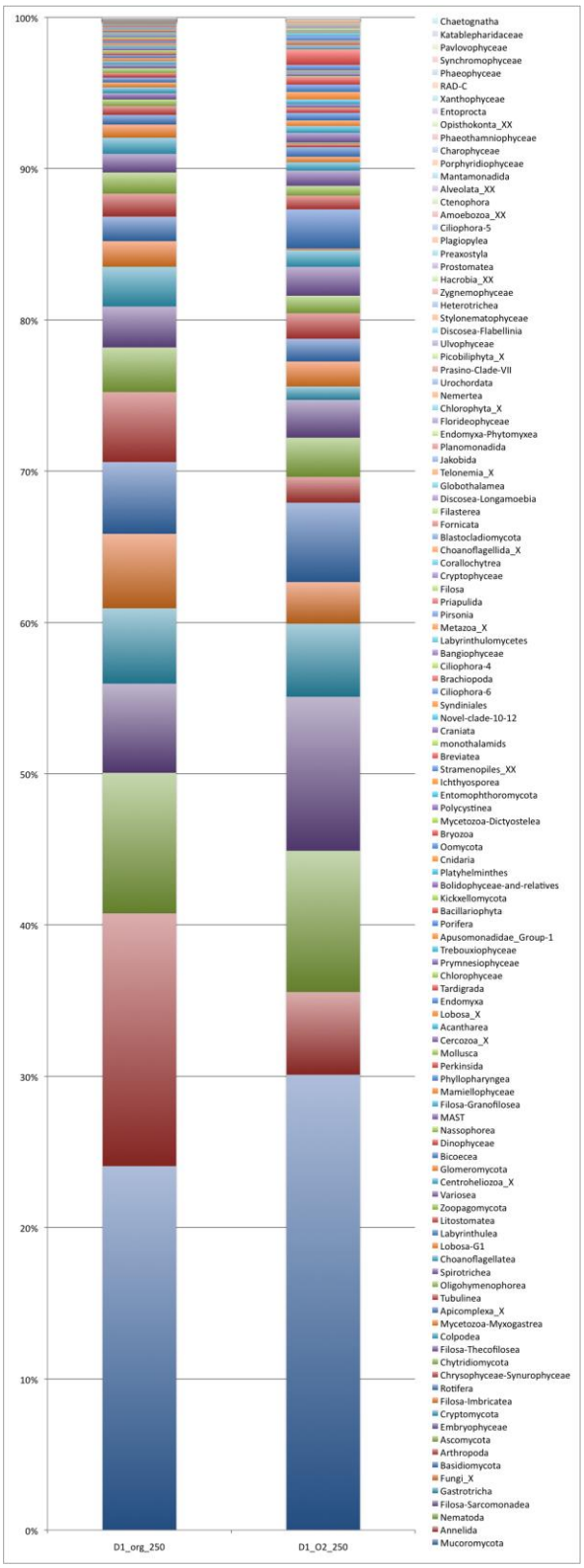

Figure 3. Biodiversity of D1 samples.
The taxonomic affiliation of all sequences in the D1-org and D1-O2 samples are shown here.

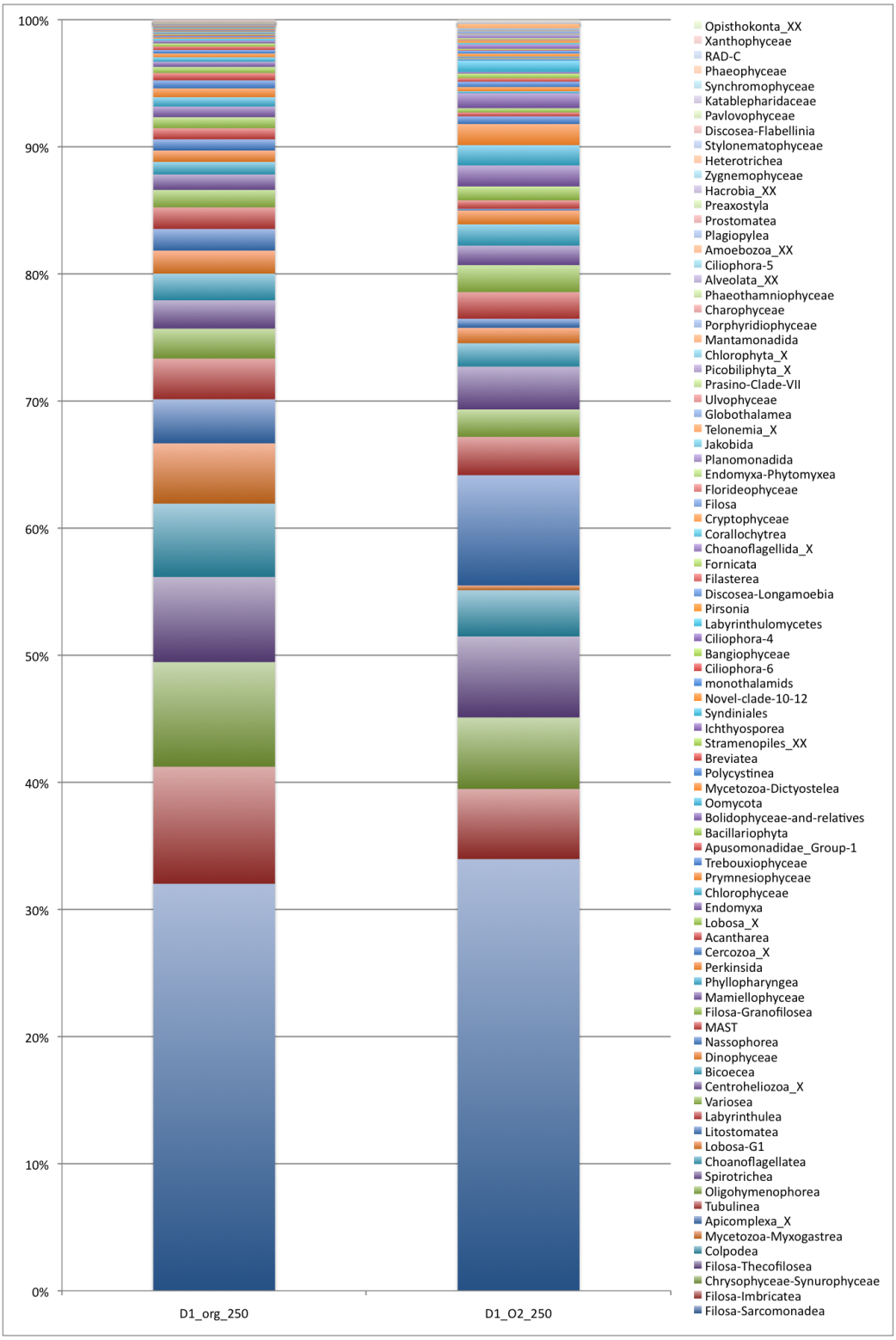

Figure 4. Biodiversity of the samples, with Metazoa, Embryophyta and Fungi removed.

As seen in Figure 5, our dataset of only the fungal sequences was dominated by sequences from the Mucoromycota[11], a group of very diverse polyphyletic fungi. These accounted for between 24% and 30% of our sequence count in both samples. Additional fungi identified were Basidiomycota[12], Ascomycota[13], Chytridomycota[14] and unclassified fungal sequences.

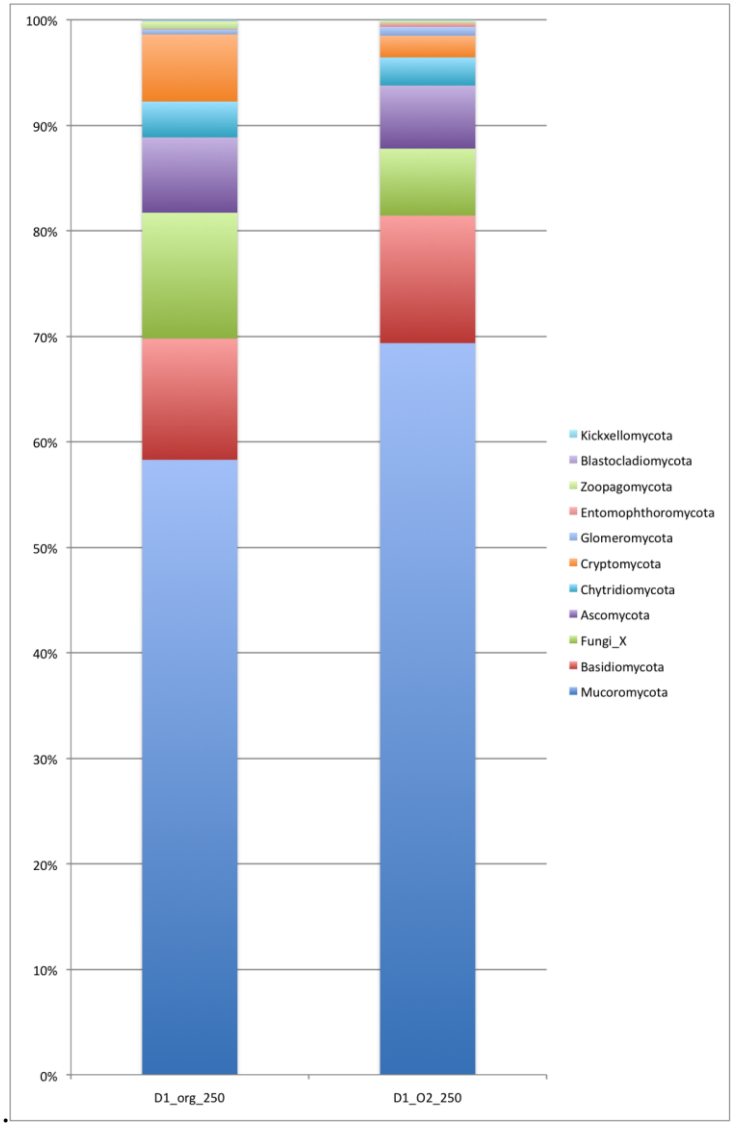

Figure 5. Biodiversity of the samples (fungi only).

[11] See http://comenius.susqu.edu/biol/202/fungi/mucoromycota/default.htm

[12] See http://www.botany.hawaii.edu/faculty/wong/Bot201/Basidiomycota/Basidiomycota.htm

[13] See https://www.princeton.edu/~achaney/tmve/wiki100k/docs/Ascomycota.html

[14] See http://www.ucmp.berkeley.edu/fungi/chytrids.html

4 DISCUSSION

In this report we have used NGS technology to assess the biodiversity of eukaryotes in samples from an undisturbed boreal forest environment in the oil sands region. This work has given us a first look at the diversity of protists and micro-invertebrates in this environment and also given us the opportunity to assess two kits or protocols of NGS workflow.

4.1 V2 2 x 250 bp vs V3 2 x 300 bp Kits

The choice of protocol is important for comparability between analyses and optimization of data collection. Given that the 2 x 300 bp kit is proposed to gather more data and longer reads, and is thus more expensive, we wanted to benchmark its performance against the less expensive 2 x 250 bp kit. Overall, the major factor that stood out was the variability between our 2 x 300 bp runs, as compared to the relative constancy of the read numbers obtained with the 2 x 250 bp kit. This discrepancy is most likely due to the use of the Qubit for DNA quantification as opposed to a specific failure of the V3 2 x 300 bp kit. The total number of reads collected in a run is proportional to the amount of sample loaded and hence the portion of the multiplexed run. Thus the $1/12^{th}$ run of the 2 x 300 bp run was not in fact $1/12^{th}$. Some libraries were more, some were less, resulting in more or less reads for a given library.

Certainly the choice of kit needs to be tailored to the question being addressed. In our case, we were looking at overall community structure and patterns and so consistency of performance between replicates is important for comparability and drawing overall conclusions. We were less concerned with identification of rare sequences. As seen in Table 1, some of the 300 runs gave much more data, and so in a study focused on novel organism discovery the trade-off between reliability and the possibility of finding new sequences might be worth using the kit. In our instance the reliability was more important.

Another major theoretical concern was that the 2 x 250 bp kit might be unable to recover the V4 region of organisms found in our sample, while the 2 x 300 bp kit would allow us to tap into this sequence diversity. This did not appear to be the case, with the 250 runs allowing for 100% contig assembly.

Thus, with these samples, the comparison of the two kits has been informative and indicates that the V2 2 x 250 bp kit is more than adequate for the analysis of diversity in our current libraries. It was used for downstream analyses and will be used in our future NGS experiments for the time being. However, changes in quality control measures integrated into the 2 x 300 bp kit protocol will need to be addressed for future runs. Once these are overcome, the 2 x 300 bp kit could well be implemented as a protocol of choice.

4.2 Biodiversity and Taxa Present

Overall, the large-scale patterns observed from our sequence count only data, showed that the most abundant taxa that were found in the samples are consistent with soil samples elsewhere. Similarly the overall biodiversity indices and species richness seemed in line with other

environments. All of these however, will have to await the results of OTU clustering, and possibly phylogenetic placement of sequences, before any quantitative assessments can be made.

We took advantage of several recent papers to do comparisons of particular taxa in our samples to those found in relatively comparable environments. A recent paper by Tedersoo et al. (2014) examined fungal biodiversity globally using an NGS approach. They sampled boreal forests in northern Eurasia (Russia, Sweden, Norway) and found their sample to be dominated by Basidiomycota, Ascomycota, Mucoromycota, Cryptomycota[15] and Chytridiomycota. By contrast our samples were clearly dominated by Mucoromycota. This may have been due to a biological sampling issue or our primer selection. However, given that we were counting unique sequences and not abundance, a single large fragment of a mucoromycete would not give this finding. On the one hand these findings do suggest an interesting biological circumstance in our soil. On the other hand, there are differences in the methodology of both sampling and analysis that count account for this discrepancy. Firstly the Tedersoo study used the ITS1-ITS2 region of the SSU rDNA gene rather than the V4 region that we used. The primers could therefore have differential bias for particular groups. Secondly, they classified their sequences using the International Nucleotide Sequence Database Consortium database, while we used the SILVA database. This would result in a different classification. Finally, their abundances were based on OTU clustered data, while ours is based on sequence count, and therefore for comparability, this step would need to be performed at the very least. Reassuringly, although the percentages differed between the studies the major groups of fungi identified as dominant were similar. The differences remain interesting and worth pursuing.

The initial goal of our work was to examine the diversity and community structure of microbial eukaryotes in this environment. We successfully identified both abundant and rare sequences in our samples, accounting for ~75 different taxonomic groups. The most relevant comparison point that we could identify, based on forest type and longitude, was a conifer forest from Arizona (Sunset Crater) examined by Bates et al. (2013). Not surprisingly the two most abundant groups found in both samples were cercozoa and ciliates. The less abundant taxa differed somewhat but were common between the Sunset Crater sample and one of our replicates. While methodologically the Bates study was more easily comparable to ours (V4 region examined, SILVA database for classification), the samples are not particularly similar and so fine-scale assessment of taxonomic distribution may not be warranted at this stage. Once OTU clustering is performed an assessment may well be worth doing.

5 CONCLUSIONS

This report provides an excellent jump-off point for further downstream analyses of protists in oil sand-associated soils. The assessment of NGS protocols gives good working guidelines for analyses of these samples using the University of Alberta TAGC facility as of December 2014. This shows that the 2 x 250 bp kit is sufficient and produced high quality, reproducible data. Furthermore, we provide a first look at the biodiversity of undisturbed boreal forest soil in the

[15] See http://en.wikipedia.org/wiki/Rozellida

Athabasca region. Although further quantitative analysis needs to be performed to accurately assess abundance and community structure, these data provide an important starting point for future assessment of reclamation and revegetation efforts. Together with the recent efforts examining microbial eukaryotic diversity within the tailings ponds (Aguilar et al. 2014, Aguilar et al. submitted), these data show the value of applying NGS approaches to better understand the natural and existing communities in various oil sands associated environments.

6 REFERENCES

Adl, M.S., 2008. Setting the tempo in land remediation: Short-term and long-term patterns in biodiversity recovery. Invited review for Microbes and Environment 23: 13-19.

Adl, M.S. and V.V.S.R. Gupta, 2006. Review: Protists in soil ecology and forest nutrient cycling. Canadian Journal of Forest Research 36: 1805-1817.

Adl, S.M., A.G. Simpson, C.E. Lane, J. Lukes, D. Bass, S.S. Bowser, M.W. Brown, F. Burki, M. Dunthorn, V. Hampl, A. Heiss, M. Hoppenrath, E. Lara, L. Le Gall, D.H. Lynn, H. McManus, E.A. Mithell, S.E. Mozley-Stanridge, L.W. Parfrey, J. Pawlowski, S. Rueckert, L. Shadwick, C.L. Schoch, A. Smirnov and F.W. Spiegel 2012. The revised classification of eukaryotes. Journal of Eukaryote Microbiology 59(5): 429-93.

Aguilar, M., E. Glücksman, D. Bass and J.B. Dacks, 2014. Next Generation Sequencing of Protists as a Measure of Microbial Community in Oil Sands Tailings Ponds: Amplicon Versus Metagenomic Approaches. Oil Sands Research and Information Network, School of Energy and the Environment, University of Alberta, Edmonton, Alberta. OSRIN Report No. TR-56. 24 pp. http://hdl.handle.net/10402/era.40100 [Last accessed December 1, 2014].

Alberta Environment, 2006. Land capability classification system for forest ecosystems in the oil sands, 3rd edition. Volume 1: Field manual for land capability determination. Prepared by the Cumulative Environmental Management Association, Fort McMurray, Alberta for Alberta Environment. Pub. No. T/875. 53 pp. plus appendices http://environment.gov.ab.ca/info/library/7707.pdf [Last accessed December 22, 2014].

Bailly, J., L. Fraissinet-Tachet, M-C. Verner, J-C. Debaud, M. Lemaire, M. Wésolowski-Louvel and R. Marmeisse, 2007. Soil eukaryotic functional diversity, a metatranscriptomic approach. ISME Journal 1: 632-642.

Bardgett, R.D., 2005. The biology of soil: A community and ecosystem approach. Oxford University Press, New York.

Bates, S.T., J.C. Clemente, G.E. Flores, W.A. Walters, L.W. Parfrey, R. Knight and N. Fierer, 2013. Global biogeography of highly diverse protistan communities in soil. ISME Journal **7**(3): 652-659.

Beckingham, J.D. and J.H. Archibald, 1996. Field guide to ecosites of Northern Alberta. Natural Resources Canada, Canadian Forest Service, Northern Forestry Centre, Edmonton, Alberta. Special Report 5.

Dimitriu, P.A. and S.J. Grayston, 2010. Relationship between soil properties and patterns of bacterial beta-diversity across reclaimed and natural boreal forest soils. Microbial Ecology 59: 563-573.

Dimitriu, P.A., C.E. Prescott, S.A. Quideau and S.J. Grayston, 2010. Impact of reclamation of surface-mined boreal forest soils on microbial community composition and function. Soil Biology and Biochemistry 42: 2289-2297.

Frouz, J., 2013. Soil biota and ecosystem development in post-mining sites. CRC Press Taylor and Francis, book in press.

Guillou, L., D. Bachar, S. Audic, D. Bass, C. Berney, L. Bittner, C. Boutte, G. Burgaud, C. de Vargas, J. Decelle, J. Del Campo, J.R. Dolan, M. Dunthorn, B. Edvardsen, M. Holzmann, W.H. Kooistra, E. Lara, N. Le Bescot, R. Logares, F. Mahé, R. Massana, M. Montresor, R. Morard, F. Not, J. Pawlowski, I. Probert, A.L. Sauvadet, R. Siano, T. Stoeck, D. Vaulot, P. Zimmermann and R. Christen, 2013. The Protist Ribosomal Reference database (PR^2): A catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. Nucleic Acids Research 41: D597-D604.

Li, W., L. Jaroszewski and A. Godzik, 2001. Clustering of highly homologous sequences to reduce the size of large protein database. Bioinformatics 17: 282-283.

Lovejoy, C., R. Massana and C. Pedrós-Alió, 2006. Diversity and distribution of marine microbial eukaryotes in the Arctic Ocean and adjacent seas. Applied and Environmental Microbiology 72(5): 3085-3095.

Magoc, T. and S. Salzberg, 2011. FLASH: Fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27: 2957-2963.

Richards, T.A., A.A. Vepritskiy, D.E. Gouliamova and S.A. Nierzwicki-Bauer, 2005. The molecular diversity of freshwater picoeukaryotes from an oligotrophic lake reveals diverse, distinctive and globally dispersed lineages. Environmental Microbiology 7(9): 1413-1425.

Sorenson, P.T., S.A. Quideau, M.D. MacKenzie, S.M. Landhäusser and S.W. Oh, 2011. Forest floor development and biochemical properties in reconstructed boreal forest soils. Applied Soil Ecology 49: 139-147.

Stoeck, T., D. Bass, M. Nebel, R. Christen, M.D. Jones, H.W. Breiner and T.A. Richards, 2010. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. Molecular Ecology 19(Suppl 1): 21-31.

Tedersoo, L., M. Bahram, S. Polme, U. Koljalg, N.S. Yorou, R. Wijesundera, L. Villarreal Ruiz, A.M. Vasco-Palacios, P.Q. Thu, A. Suija, M.E. Smith, C. Sharp, E. Saluveer, A. Saitta, M. Rosas, T. Riit, D. Ratkowsky, K. Pritsch, K. Poldmaa, M. Piepenbring, C. Phosri, M. Peterson, K. Parts, K. Partel, E. Otsing, E. Nouhra, A.L. Njouonkou, R.H. Nilsson, L.N. Morgado, J. Mayor, T.W. May, L. Majuakim, D.J. Lodge, S.S. Lee, K.H. Larsson, P. Kohout, K. Hosaka, I. Hiiesalu, T.W. Henkel, H. Harend, L.D. Guo, A. Greslebin, G. Grelet, J. Geml, G. Gates, W. Dunstan, C. Dunk, R. Drenkhan, J. Dearnaley, A. De Kesel, T. Dang,

X. Chen, F. Buegger, F.Q. Brearley, G. Bonito, S. Anslan, S. Abell and K. Abarenkov, 2014. Fungal biogeography. Global diversity and geography of soil fungi. Science 346(6213): 1256688.

Turcotte, I., S.A. Quideau and S.W. Oh, 2009. Organic matter quality in reclaimed boreal forest soils following oilsands mining. Organic Geochemistry 40: 510-519.

7 GLOSSARY

7.1 Terms

Amplicon

Region of DNA produced via an amplification reaction, in this case by a polymerase chain reaction.

Base Pair

DNA is assembled of polymerized deoxyribonucleic acids (referred to as a base). Each DNA chain is composed of two polymers whereby the corresponding subunits interact or pair-up via hydrogen bonds. The unit of two bases that are bonded, one on each chain, is deemed a base pair.

Bioinformatic

The use of computational methods to analyze biological information, in this case DNA sequence. Often this involves large-scale analyses of data and may involve programming or modification of scripts for use on computational clusters, rather than the use of 'out of the box' software.

Chimaeric Sequence

In the sequencing process, using a "paired end read" approach, the same DNA molecule is read from both ends (5' and 3'). Therefore an important step of the post-processing bioinformatic analysis is pairing up of sequences that the correct reads from both ends of the same DNA molecule. In cases where the program incorrectly assembles reads from two different molecules and treats it as a new or unique sequence, this is deemed a chimaeric sequence.

Contigs

A set of overlapping DNA segments that together represent a consensus region of DNA.

Eukaryote

One of the three fundamental "Domains" of life, as assessed by phylogenetic methods that judge genetic distance and by a suite or shared morphological features that are exclusive to these organisms. Although defined by the presence of genomic DNA enclosed by a double lipid bilayer (the nuclear envelope/endoplasmic reticulum), the presence of other organelles such as the Golgi body, mitochondria-related organelle, endosomes are also often taken as pan-eukaryotic features.

Homopolymer

Any substance that is composed of a single repeating subunit. In this case, this refers to regions of DNA that have the same single DNA base in succession.

Library

A given sample of DNA with the relevant oligonucleotide adaptors added to each DNA strand, allowing it to be analyzed by the sequencing platform.

Next Generation Sequencing

Refers to a set of post-sanger sequencing methodologies that produce large numbers of, often short, reads. Also treated as synonymous with "high through-put" sequencing. In this case we use the Illumina platform.

Operational Taxonomic Unit (OTU)

This term describes a proxy for "species-level" identity in DNA sequences derived from the environment, where the organisms from which the DNA was taken were not identified. An OTU made up of numerous sequence "reads" would be regarded as representing numerous instances of the organism from which the DNA sequence came. However in cases where an organism has more than one copy of the DNA sequence, there may not be a linear 1:1 match between number of reads and number of organisms.

Paired End

Protocol for next generation sequencing where the same molecule is sequenced from both directions and the reads are matched.

Richness

The total number of taxonomic groups present in a community.

Sequence

The order of bases in a DNA chain or molecule.

Shannon Index

A measure of diversity that simultaneously takes into account how many taxonomical groups are present and how evenly distributed the organisms are across the groups. The index increases both with an increasing number of groups and with a more uniform distribution of the organisms between the groups. It is a measure of entropy in the system, as the higher the number of groups and the more equally distributed the individuals between the different groups the more difficult is to predict to which group one randomly selected individual belongs to.

The formula used for its calculation is:

$$H' = -\sum_{i=1}^{R} p_i \ln p_i$$

Where R is the number of groups and p is the number of individuals that belong to each group.

Simpson Index

A measure of diversity based on similar principles as the Shannon index. It is calculated using the formula below:

$$\lambda = \sum_{i=1}^{R} p_i^2$$

Where R represents the number of groups and p is the number of individuals that belong to each group.

Uchime Package

A computational package used, in our analyses, for detecting potential chimaeras. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3150044/

7.2 Acronyms

bp	Base Pairs
DNA	Deoxyribonucleic Acid
NGS	Next Generation Sequencing
OSRIN	Oil Sands Research and Information Network
OTU	Operational Taxonomic Unit
PCR	Polymerase Chain Reaction
SEE	School of Energy and the Environment

LIST OF OSRIN REPORTS

OSRIN reports are available on the University of Alberta's Education & Research Archive at http://hdl.handle.net/10402/era.17209. The Technical Report (TR) series documents results of OSRIN funded projects. The Staff Reports (SR) series represent work done by OSRIN staff.

OSRIN Technical Reports – http://hdl.handle.net/10402/era.17507

BGC Engineering Inc., 2010. Oil Sands Tailings Technology Review. OSRIN Report No. TR-1. 136 pp. http://hdl.handle.net/10402/era.17555

BGC Engineering Inc., 2010. Review of Reclamation Options for Oil Sands Tailings Substrates. OSRIN Report No. TR-2. 59 pp. http://hdl.handle.net/10402/era.17547

Chapman, K.J. and S.B. Das, 2010. Survey of Albertans' Value Drivers Regarding Oil Sands Development and Reclamation. OSRIN Report TR-3. 13 pp. http://hdl.handle.net/10402/era.17584

Jones, R.K. and D. Forrest, 2010. Oil Sands Mining Reclamation Challenge Dialogue – Report and Appendices. OSRIN Report No. TR-4. 258 pp. http://hdl.handle.net/10402/era.19092

Jones, R.K. and D. Forrest, 2010. Oil Sands Mining Reclamation Challenge Dialogue – Report. OSRIN Report No. TR-4A. 18 pp. http://hdl.handle.net/10402/era.19091

James, D.R. and T. Vold, 2010. Establishing a World Class Public Information and Reporting System for Ecosystems in the Oil Sands Region – Report and Appendices. OSRIN Report No. TR-5. 189 pp. http://hdl.handle.net/10402/era.19093

James, D.R. and T. Vold, 2010. Establishing a World Class Public Information and Reporting System for Ecosystems in the Oil Sands Region – Report. OSRIN Report No. TR-5A. 31 pp. http://hdl.handle.net/10402/era.19094

Lott, E.O. and R.K. Jones, 2010. Review of Four Major Environmental Effects Monitoring Programs in the Oil Sands Region. OSRIN Report No. TR-6. 114 pp. http://hdl.handle.net/10402/65.20287

Godwalt, C., P. Kotecha and C. Aumann, 2010. Oil Sands Tailings Management Project. OSRIN Report No. TR-7. 64 pp. http://hdl.handle.net/10402/era.22536

Welham, C., 2010. Oil Sands Terrestrial Habitat and Risk Modeling for Disturbance and Reclamation – Phase I Report. OSRIN Report No. TR-8. 109 pp. http://hdl.handle.net/10402/era.22567

Schneider, T., 2011. Accounting for Environmental Liabilities under International Financial Reporting Standards. OSRIN Report TR-9. 16 pp. http://hdl.handle.net/10402/era.22741

Davies, J. and B. Eaton, 2011. Community Level Physiological Profiling for Monitoring Oil Sands Impacts. OSRIN Report No. TR-10. 44 pp. http://hdl.handle.net/10402/era.22781

Hurndall, B.J., N.R. Morgenstern, A. Kupper and J. Sobkowicz, 2011. Report and Recommendations of the Task Force on Tree and Shrub Planting on Active Oil Sands Tailings Dams. OSRIN Report No. TR-11. 15 pp. http://hdl.handle.net/10402/era.22782

Gibson, J.J., S.J. Birks, M. Moncur, Y. Yi, K. Tattrie, S. Jasechko, K. Richardson, and P. Eby, 2011. Isotopic and Geochemical Tracers for Fingerprinting Process-Affected Waters in the Oil Sands Industry: A Pilot Study. OSRIN Report No. TR-12. 109 pp. http://hdl.handle.net/10402/era.23000

Oil Sands Research and Information Network, 2011. Equivalent Land Capability Workshop Summary Notes. OSRIN Report TR-13. 83 pp. http://hdl.handle.net/10402/era.23385

Kindzierski, W., J. Jin and M. Gamal El-Din, 2011. Plain Language Explanation of Human Health Risk Assessment. OSRIN Report TR-14. 37 pp. http://hdl.handle.net/10402/era.23487

Welham, C. and B. Seely, 2011. Oil Sands Terrestrial Habitat and Risk Modelling for Disturbance and Reclamation – Phase II Report. OSRIN Report No. TR-15. 93 pp. http://hdl.handle.net/10402/era.24547

Morton Sr., M., A. Mullick, J. Nelson and W. Thornton, 2011. Factors to Consider in Estimating Oil Sands Plant Decommissioning Costs. OSRIN Report No. TR-16. 62 pp. http://hdl.handle.net/10402/era.24630

Paskey, J. and G. Steward, 2012. The Alberta Oil Sands, Journalists, and Their Sources. OSRIN Report No. TR-17. 33 pp. http://hdl.handle.net/10402/era.25266

Cruz-Martinez, L. and J.E.G. Smits, 2012. Potential to Use Animals as Monitors of Ecosystem Health in the Oil Sands Region – July 2013 Update. OSRIN Report No. TR-18. 59 pp. http://hdl.handle.net/10402/era.25417

Hashisho, Z., C.C. Small and G. Morshed, 2012. Review of Technologies for the Characterization and Monitoring of VOCs, Reduced Sulphur Compounds and CH_4. OSRIN Report No. TR-19. 93 pp. http://hdl.handle.net/10402/era.25522

Kindzierski, W., J. Jin and M. Gamal El-Din, 2012. Review of Health Effects of Naphthenic Acids: Data Gaps and Implications for Understanding Human Health Risk. OSRIN Report No. TR-20. 43 pp. http://hdl.handle.net/10402/era.26060

Zhao, B., R. Currie and H. Mian, 2012. Catalogue of Analytical Methods for Naphthenic Acids Related to Oil Sands Operations. OSRIN Report No. TR-21. 65 pp. http://hdl.handle.net/10402/era.26792

Oil Sands Research and Information Network and Canadian Environmental Assessment Agency, 2012. Summary of the Oil Sands Groundwater – Surface Water Interactions Workshop. OSRIN Report No. TR-22. 125 pp. http://hdl.handle.net/10402/era.26831

Valera, E. and C.B. Powter, 2012. Implications of Changing Environmental Requirements on Oil Sands Royalties. OSRIN Report No. TR-23. 21 pp. http://hdl.handle.net/10402/era.27344

Dixon, R., M. Maier, A. Sandilya and T. Schneider, 2012. Qualifying Environmental Trusts as Financial Security for Oil Sands Reclamation Liabilities. OSRIN Report No. TR-24. 32 pp. http://hdl.handle.net/10402/era.28305

Creasey, R., 2012. Professional Judgment in Mineable Oil Sands Reclamation Certification: Workshop Summary. OSRIN Report No. TR-25. 52 pp. http://hdl.handle.net/10402/era.28331

Alberta Innovates – Technology Futures, 2012. Investigating a Knowledge Exchange Network for the Reclamation Community. OSRIN Report No. TR-26. 42 pp. http://hdl.handle.net/10402/era.28407

Dixon, R.J., J. Kenney and A.C. Sandilya, 2012. Audit Protocol for the Mine Financial Security Program. OSRIN Report No. TR-27. 27 pp. http://hdl.handle.net/10402/era.28514

Davies, J., B. Eaton and D. Humphries, 2012. Microcosm Evaluation of Community Level Physiological Profiling in Oil Sands Process Affected Water. OSRIN Report No. TR-28. 33 pp. http://hdl.handle.net/10402/era.29322

Thibault, B., 2012. Assessing Corporate Certification as Impetus for Accurate Reporting in Self-Reported Financial Estimates Underlying Alberta's Mine Financial Security Program. OSRIN Report No. TR-29. 37 pp. http://hdl.handle.net/10402/era.29361

Pyper, M.P., C.B. Powter and T. Vinge, 2013. Summary of Resiliency of Reclaimed Boreal Forest Landscapes Seminar. OSRIN Report No. TR-30. 131 pp. http://hdl.handle.net/10402/era.30360

Pyper, M. and T. Vinge, 2013. A Visual Guide to Handling Woody Materials for Forested Land Reclamation. OSRIN Report No. TR-31. 10 pp. http://hdl.handle.net/10402/era.30381

Mian, H., N. Fassina, A. Mukherjee, A. Fair and C.B. Powter, 2013. Summary of 2013 Tailings Technology Development and Commercialization Workshop. OSRIN Report No. TR-32. 69 pp. http://hdl.handle.net/10402/era.31012

Howlett, M. and J. Craft, 2013. Application of Federal Legislation to Alberta's Mineable Oil Sands. OSRIN Report No. TR-33. 94 pp. http://hdl.handle.net/10402/era.31627

Welham, C., 2013. Factors Affecting Ecological Resilience of Reclaimed Oil Sands Uplands. OSRIN Report No. TR-34. 44 pp. http://hdl.handle.net/10402/era.31714

Naeth, M.A., S.R. Wilkinson, D.D. Mackenzie, H.A. Archibald and C.B. Powter, 2013. Potential of LFH Mineral Soil Mixes for Land Reclamation in Alberta. OSRIN Report No. TR-35. 64 pp. http://hdl.handle.net/10402/era.31855

Welham, C. and B. Seely, 2013. Oil Sands Terrestrial Habitat and Risk Modelling for Disturbance and Reclamation: The Impact of Climate Change on Tree Regeneration and Productivity – Phase III Report. OSRIN Report No. TR-36. 65 pp. http://hdl.handle.net/10402/era.31900

Eaton, B., T. Muhly, J. Fisher and S-L. Chai, 2013. Potential Impacts of Beaver on Oil Sands Reclamation Success – an Analysis of Available Literature. OSRIN Report No. TR-37. 65 pp. http://hdl.handle.net/10402/era.32764

Paskey, J., G. Steward and A. Williams, 2013. The Alberta Oil Sands Then and Now: An Investigation of the Economic, Environmental and Social Discourses Across Four Decades. OSRIN Report No. TR-38. 108 pp. http://hdl.handle.net/10402/era.32845

Watson, B.M. and G. Putz, 2013. Preliminary Watershed Hydrology Model for Reclaimed Oil Sands Sites. OSRIN Report No. TR-39. 193 pp. http://hdl.handle.net/10402/era.34250

Birks, S.J., Y. Yi, S. Cho, J.J. Gibson and R. Hazewinkel, 2013. Characterizing the Organic Composition of Snow and Surface Water in the Athabasca Region. OSRIN Report No. TR-40. 62 pp. http://hdl.handle.net/10402/era.36643

De Corby, R.G., 2013. Development of Silicon-Based Optofluidic Sensors for Oil Sands Environmental Monitoring. OSRIN Report No. TR-41. 19 pp. http://hdl.handle.net/10402/era.36936

Iqbal, M., T.K. Purkait, J.G.C. Veinot and G.G. Goss, 2013. Benign-by-Design: Synthesis of Engineered Silicon Nanoparticles and their Application to Oil Sands Water Contaminant Remediation. OSRIN Report No. TR-42. 30 pp. http://hdl.handle.net/10402/era.37308

Oil Sands Research and Information Network, 2013. Future of Shrubs in Oil Sands Reclamation Workshop. OSRIN Report No. TR-43. 71 pp. http://hdl.handle.net/10402/era.37440

Smreciu, A., K. Gould and S. Wood, 2013. Boreal Plant Species for Reclamation of Athabasca Oil Sands Disturbances – Updated December 2014. OSRIN Report No. TR-44. 23 pp. plus appendices. http://hdl.handle.net/10402/era.37533

Pereira, A.S. and J.W. Martin, 2014. On-Line Solid Phase Extraction – HPLC – Orbitrap Mass Spectrometry for Screening and Quantifying Targeted and Non-Targeted Analytes in Oil Sands Process-Affected Water and Natural Waters in the Athabasca Oil Sands Region. OSRIN Report No. TR-45. 33 pp. http://hdl.handle.net/10402/era.37793

Liang, J., F. Tumpa, L.P. Estrada, M. Gamal El-Din and Y. Liu, 2014. Ozone-Assisted Settling of Diluted Oil Sands Mature Fine Tailings: A Mechanistic Study. OSRIN Report No. TR-46. 43 pp. http://hdl.handle.net/10402/era.38226

Rochdi, N., J. Zhang, K. Staenz, X. Yang, D. Rolfson, J. Banting, C. King and R. Doherty, 2014. Monitoring Procedures for Wellsite, In-Situ Oil Sands and Coal Mine Reclamation in Alberta. OSRIN Report No. TR-47. 156 pp. http://hdl.handle.net/10402/era.38742

Taheriazad, L., C. Portillo-Quintero and G.A. Sanchez-Azofeifa, 2014. Application of Wireless Sensor Networks (WSNs) to Oil Sands Environmental Monitoring. OSRIN Report No. TR-48. 51 pp. http://hdl.handle.net/10402/era.38858

Marey, H.S., Z. Hashisho and L. Fu, 2014. Satellite Remote Sensing of Air Quality in the Oil Sands Region. OSRIN Report No. TR-49. 104 pp. http://hdl.handle.net/10402/era.38882

Li, C., A. Singh, N. Klamerth, K. McPhedran, P. Chelme-Ayala, M. Belosevic and M. Gamal El-Din, 2014. Synthesis of Toxicological Behavior of Oil Sands Process-Affected Water Constituents. OSRIN Report No. TR-50. 101 pp. http://hdl.handle.net/10402/era.39659

Jiang, Y. and Y. Liu, 2014. Application of Forward Osmosis Membrane Technology for Oil Sands Process-Affected Water Desalination. OSRIN Report No. TR-51. 27 pp. http://hdl.handle.net/10402/era.39855

Zhu, L., M. Yu, L. Delgado Chávez, A. Ulrich and T. Yu, 2014. Review of Bioreactor Designs Applicable to Oil Sands Process-Affected Water Treatment. OSRIN Report No. TR-52. 39 pp. http://hdl.handle.net/10402/era.39903

Oil Sands Research and Information Network, 2014. Oil Sands Rules, Tools and Capacity: Are we Ready for Upcoming Challenges? OSRIN Report No. TR-53. 120 pp. http://hdl.handle.net/10402/era.39985

Iqbal, M., T.K. Purkait, M. Aghajamali, L. Hadidi, J.G.C. Veinot, G.G. Goss and M. Gamal El-Din, 2014. Hybrid Aerogel SiNP Membranes for Photocatalytic Remediation of Oil Sands Process Water. OSRIN Report No. TR-54. 29 pp. http://hdl.handle.net/10402/era.40004

Schoonmaker, A., J-M. Sobze, E. Fraser, E. Marenholtz, A. Smreciu, C.B. Powter and M. Mckenzie, 2014. Alternative Native Boreal Seed and Plant Delivery Systems for Oil Sands Reclamation. OSRIN Report No. TR-55. 61 pp. http://hdl.handle.net/10402/era.40099

Aguilar, M., E. Glücksman, D. Bass and J.B. Dacks, 2014. Next Generation Sequencing of Protists as a Measure of Microbial Community in Oil Sands Tailings Ponds: Amplicon Versus Metagenomic Approaches. OSRIN Report No. TR-56. 24 pp. http://hdl.handle.net/10402/era.40100

Alessi, D.S., M.S. Alam and M.C. Kohler, 2014. Designer Biochar-Coke Mixtures to Remove Naphthenic Acids from Oil Sands Process-Affected Water (OSPW). OSRIN Report No. TR-57. 38 pp. http://hdl.handle.net/10402/era.40122

Oil Sands Research and Information Network, 2014. Survey of Oil Sands Environmental Management Research and Information Needs. OSRIN Report No. TR-58. 67 pp. http://hdl.handle.net/10402/era.40128

Huang, Q., H. Wang and M.A. Lewis, 2014. Development of a Toxin-Mediated Predator-Prey Model Applicable to Aquatic Environments in the Athabasca Oil Sands Region. OSRIN Report No. TR-59. 59 pp. http://hdl.handle.net/10402/era.40140

Currie, R., S. Bansal, I. Khan and H. Mian, 2014. An Investigation of the Methylene Blue Titration Method for Clay Activity of Oil Sands Samples. OSRIN Report No. TR-60. 50 pp. http://hdl.handle.net/10402/era.40164

Welham, C., 2014. Risk and Uncertainty in Oil Sands Upland Reclamation: Best Management Practices within the Context of Climate Change. OSRIN Report No. TR-61. 26 pp. http://hdl.handle.net/10402/era.40171

Mahdavi, H., H. Mian, S. Hepperle and Z. Burkus, 2014. Standard Operating Procedures for Analysis of Naphthenic Acids from Oil Sands Process-Affected Water. OSRIN Report No. TR-62. 67 pp. http://hdl.handle.net/10402/era.40181

McPhedran, K., M.S. Islam and M. Gamal El-Din, 2014. Development of a Novel Engineered Bioprocess for Oil Sands Process-Affected Water and Tailings Fines/Bitumen/Water Separation. OSRIN Report No. TR-63. 28 pp. http://hdl.handle.net/10402/era.40190

Birks, J., Y. Yi, S. Cho, E. Taylor and J. Gibson, 2014. Characterizing the Organic Composition of Snow and Surface Water Across the Athabasca Region: Phase 2. OSRIN Report No. TR-64. 47 pp. http://hdl.handle.net/10402/era.40243

Alberta Centre for Reclamation and Restoration Ecology and Oil Sands Research and Information Network, 2014. Creating a Knowledge Platform for the Reclamation and Restoration Ecology Community: Expanding the OSRIN Model Beyond the Oil Sands. OSRIN Report No. TR-65. 19 pp. http://hdl.handle.net/10402/era.40323

Liang, J., Z. Guo, L. Deng and Y. Liu, 2014. MFT Consolidation Through Microbial Induced Calcium Carbonate Precipitation. OSRIN Report No. TR-66. 31 pp. http://hdl.handle.net/10402/era.40330

Eaton, B.R., J.T. Fisher, G.T. McKenna, and J. Pollard. 2014. An Ecological Framework for Wildlife Habitat Design for Oil Sands Mine Reclamation. OSRIN Report No. TR-67. 83 pp. http://hdl.handle.net/10402/era.40338

Hopkins, D, K. Wall and C. Wilson, 2014. Measured Concentrations of Metals and Polycyclic Aromatic Hydrocarbons in Plants, Berries and Soil Located North of Fort McMurray, Alberta. OSRIN Report No. TR-68. 134 pp. http://hdl.handle.net/10402/era.40339

OSRIN Videos – http://hdl.handle.net/10402/era.29304

Rooney Productions, 2012. Assessment Methods for Oil Sands Reclamation Marshes. OSRIN Video No. V-1. 20 minutes. Also available on the University of Alberta You Tube Channel (recommended approach).

Rooney Productions, 2012. Assessment Methods for Oil Sands Reclamation Marshes. OSRIN Video No. V-1. Nine-part mobile device version. Also available on the University of Alberta You Tube Channel (link to Part 1 - recommended approach).

OSRIN Staff Reports – http://hdl.handle.net/10402/era.19095

OSRIN, 2010. Glossary of Terms and Acronyms used in Oil Sands Mining, Processing and Environmental Management – December 2014 Update. OSRIN Report No. SR-1. 125 pp. http://hdl.handle.net/10402/era.17544

OSRIN, 2010. OSRIN Writer's Style Guide – November 2013 Update. OSRIN Report No. SR-2. 29 pp. http://hdl.handle.net/10402/era.17545

OSRIN, 2010. OSRIN Annual Report: 2009/2010. OSRIN Report No. SR-3. 27 pp. http://hdl.handle.net/10402/era.17546

OSRIN, 2010. Guide to OSRIN Research Grants and Services Agreements - June 2011 Update. OSRIN Report No. SR-4. 21 pp. http://hdl.handle.net/10402/era.17558

OSRIN, 2011. Summary of OSRIN Projects – October 2014 Update. OSRIN Report No. SR-5. 113 pp. http://hdl.handle.net/10402/era.20529

OSRIN, 2011. OSRIN Annual Report: 2010/11. OSRIN Report No. SR-6. 34 pp. http://hdl.handle.net/10402/era.23032

OSRIN, 2011. OSRIN's Design and Implementation Strategy. OSRIN Report No. SR-7. 10 pp. http://hdl.handle.net/10402/era.23574

OSRIN, 2012. OSRIN Annual Report: 2011/12. OSRIN Report No. SR-8. 25 pp. http://hdl.handle.net/10402/era.26715

OSRIN, 2013. OSRIN Annual Report: 2012/13. OSRIN Report No. SR-9. 56 pp. http://hdl.handle.net/10402/era.31211

OSRIN, 2014. OSRIN Annual Report: 2013/14. OSRIN Report No. SR-10. 66 pp. http://hdl.handle.net/10402/era.38508

OSRIN, 2014. OSRIN's Did You Know Series: The Collected Works. OSRIN Report No. SR-11. 163 pp. http://hdl.handle.net/10402/era.40220

OSRIN, 2014. Media Coverage of Oil Sands Pipelines: A Chronological Record of Headlines from 2010 to 2014. OSRIN Report No. SR-12. 140 pp. http://hdl.handle.net/10402/era.40331