HHS Public Access

Author manuscript

J Invest Dermatol. Author manuscript; available in PMC 2015 February 01.

Published in final edited form as:

J Invest Dermatol. 2014 August ; 134(8): 2138–2145. doi:10.1038/jid.2014.127.

High *Rhodotorula* sequences in skin transcriptome of patients with diffuse systemic sclerosis

Sarah T. Arron[1,‡], **Michelle T. Dimon**[1], **Zhenghui Li**, **Michael E. Johnson**[2], **Tammara Wood**, **Luzviminda Feeney**[1], **Jorge Gil Angeles**[1], **Robert Lafyatis**[3], and **Michael L. Whitfield**[2,‡]

[1]Department of Dermatology, University of California, San Francisco, San Francisco, CA

[2]Department of Genetics, Dartmouth Geisel School of Medicine, Hanover, NH

[3]Department of Medicine, Boston University School of Medicine, Boston, MA

Abstract

Previous studies have suggested a role for pathogens as a trigger of systemic sclerosis (SSc), though neither a pathogen nor a mechanism of pathogenesis is known. Here we show enrichment of *Rhodotorula* sequences in the skin of patients with early, diffuse SSc compared to normal controls. RNA-seq was performed on four SSc and four controls, to a depth of 200 million reads per patient. Data were analyzed to quantify the non-human sequence reads in each sample. We found little difference between bacterial microbiome and viral read counts, but found a significant difference between the read counts for a mycobiome component, *R. glutinis*. Normal samples contained almost no detected *R. glutinis* or other *Rhodotorula* sequence reads (mean score 0.021 for *R. glutinis*, 0.024 for all *Rhodotorula*). In contrast, SSc samples had a mean score of 5.039 for *R. glutinis* (5.232 for *Rhodotorula*). We were able to assemble the D1–D2 hypervariable region of the 28S rRNA of *R. glutinis* from each of the SSc samples. Taken together, these results suggest *R. glutinis* may be present in the skin of early SSc patients at higher levels than normal skin, raising the possibility that it may be triggering the inflammatory response found in SSc.

Introduction

Systemic Sclerosis (SSc) is a rare and poorly understood systemic autoimmune disease that results in skin fibrosis and severe internal organ involvement. There is a limited understanding of its pathophysiology and there is little data to indicate what may trigger the disease. One in three patients dies within 10 years of diagnosis (Steen and Medsger 2007); there are no validated diagnostic markers and no curative treatments.

‡To whom correspondence should be addressed: Sarah T. Arron, Department of Dermatology, University of California, San Francisco, 1701 Divisadero Street, Box 0316, San Francisco, CA 94143-0316, tel: (415) 353-7839, fax: (415) 353-7838, arrons@derm.ucsf.edu and Michael L. Whitfield, Ph.D., Department of Genetics, Geisel School of Medicine at Dartmouth, 7400 Remsen, Hanover, NH 03755, phone 603.650.1109, fax 603.650.1188, michael.L.whitfield@dartmouth.edu.

Conflict of Interest

Dr. Whitfield has filed patents for gene expression biomarkers in systemic sclerosis and is a scientific founder of Celdara Medical LLC. The remaining authors declare no conflict of interest.

We have demonstrated gene expression based subsets within SSc patients (Whitfield et al. 2003; Milano et al. 2008; Chung et al. 2009; Sargent et al. 2009; Pendergrass et al. 2012) by analysis of skin biopsies in three independent cohorts (Milano et al. 2008; Pendergrass et al. 2012; Hinchcliff et al. 2013). Using genome-wide and bioinformatic-driven strategies (Sargent et al. 2009; Greenblatt et al. 2012), SSc patients may now be divided into pathway-centric subsets. These are the inflammatory, fibroproliferative, limited and normal-like subsets (Milano et al. 2008; Pendergrass et al. 2012; Hinchcliff et al. 2013). The inflammatory subset of patients is characterized by infiltrating immune cells that include T and B lymphocytes and macrophages (Milano et al. 2008; Greenblatt et al. 2012; Pendergrass et al. 2012; Hinchcliff et al. 2013). We have recently shown that two major pathways driving fibrosis in the inflammatory subset of patients are the profibrotic IL-13 and IL-4 pathways, which signal through a shared receptor IL-4RA (Greenblatt et al. 2012). We have also demonstrated that SSc patients that map to the inflammatory subset show improvement while taking a commonly used SSc therapeutic agent, mycophenolate mofetile (MMF), while the patients in the fibroproliferative subset do not show any clinical improvement (Hinchcliff et al. 2013). We have shown that the gene expression subsets are stable over periods of 6 – 12 months (Pendergrass et al. 2012) although recent meta-analysis of all published datasets suggest the groups may be long lived, but interconnected (Mahoney, Johnson, Whitfield, *Submitted*). In such a longitudinal model, the inflammatory group may be a key point in the initiation of disease since most genetic changes that have been associated with SSc risk occur in the immune system, suggesting that genetics along with some environmental trigger is an initiating event in SSc.

Identification of such environment triggers for most systemic autoimmune diseases has been elusive despite the significant health burden these diseases impose worldwide. Links have been suggested between SSc and cytomegalovirus, parvovirus B19, Epstein-Barr virus, endogenous retroviruses, and Chlamydia (Hamamdzic et al. 2002; Grossman et al. 2011), but these reports have not been substantiated. Hypotheses include molecular mimicry, in which homology between pathogen and self-peptides results in cross-activation of autoreactive lymphocytes; chronic inflammation and endothelial cell damage; and microbial superantigens activating immune response in the absence of cognate antigen (Grossman et al. 2011).

High-throughput sequencing technologies allow analysis of both host and pathogen-expressed sequences and genomes to identify exogenous viral, bacterial or fungal triggers of disease (the metagenome). Metagenomic analyses allow an unbiased assessment of all microorganisms in a complex disease sample. Here we present a comprehensive characterization of the metagenome in the skin of early, active SSc patients that show an inflammatory gene expression signature. Analysis of the full metagenome (including the microbiome, fungal mycobiome, and viral sequences) in patients with these earliest signs of disease, and mapping host and metagenomic sequences has identified a common environmental fungus, *Rhodotorula glutinis* as over-represented in SSc skin. Our preliminary studies suggest that disease pathogenesis may include a common environmental trigger that we hypothesize elicits immune activation in a permissive host genetic background.

Results

Patient characteristics

Lesional skin was obtained from the forearm of four patients with early, diffuse SSc within 6 months of first onset of non-Raynaud's symptoms (Table 1). All patients were in the inflammatory intrinsic subset. Two patients were untreated and two had received low-dose immunosuppression. Control skin was obtained from the forearm of healthy patients.

Initial metagenomic analysis shows fungal reads in SSc samples

We performed quality filtering and human sequence filtering using the human genome (hg19). Over 99% of the total readset was derived from human or nonhuman primate in both SSc and control samples. On average, 4×10^5 reads remained per sample after host filtering (Table 2). IMSA mapped reads to the NCBI non-redundant nucleotide (nt) database and generated taxonomy reports. In this analysis, each taxonomic level is given a score based on the number of reads aligning to sequences in that taxonomic category, where reads with multiple best alignments generate partial scores for each category with an alignment. Figure 1 demonstrates the breakdown of read scores in the dataset by taxonomic division. The microbial reads had a significantly different distribution between SSc and normal samples, with significantly more reads in SSc samples mapping to the plant and fungal division. At the top level, IMSA uses the GenBank divisions for an overview of metagenomic results. In this organization, plant and fungal sequences are combined (Ouellette and Boguski 1997). There were no significant differences in the scores derived from bacteria or viruses between SSc and control.

Figure 2 shows a heat map of taxonomy scores for bacterial, viral and fungal genera. For this analysis and all subsequent analyses, only reads with a single best alignment were retained to prevent noise from reads aligning across multiple species. Unsupervised clustering discriminated between normal and SSc samples, with the cluster driven by fungal genera in the *Basidiomycota* phylum (Figure 2). Common skin colonizers *Streptococcus*, *Propionibacterium* and *Malassezia* were represented across all samples, while SSc samples were enriched in genera of order *Sporidiobolales*, including *Rhodotorula*, *Rhodosporidium*, and *Sporobolomyces*.

SSc samples contain significantly more reads derived from Rhodotorula species

To determine the source of these differences between SSc and normal samples, TaxMaps were generated to visualize the scores of the taxonomic categories inside the plant and fungal division (Figure 3). Average TaxMaps for the normal samples showed *Malassezia globosa* and *Trimorphomyces papilionaceus* as the only species with an average normalized score above 0.05 (Figure 3A). By contrast, the SSc samples had more diverse fungal sequences, with *M. globosa*, *T. papilionaceus* but also *Bullera sakaeratica*, *Leucosporidium sp AY30*, *Rhodotorula hordea*, *Rhodotorula glutinis* and *Rhodotorula mucilaginosa* all showing average scores above 0.05 (supplemental information contains each individual TaxMap with all scores above 0.01). The most striking difference between SSc and normal TaxMaps is the large number of *R. glutinis* and *R. mucilaginosa* reads (Figure 3B). Normal samples averaged a total score of 0.55 for the entire plant and fungal division with no

species having an average score over 0.11. In the SSc samples, *R. glutinis* had an average score of 5.04, while the closely related *R. mucilaginosa* had an average score of 0.20.

SSc samples had a 252-fold increase in *R. glutinis* score per million total reads (normal mean score= 0.021, 95% CI −0.01–0.05, SSc mean score 5.039, 95% CI 2.97–7.11, Wilcoxon rank-sum p=0.01) (Figure 3C).

Assembly of *Rhodotorula* contigs indicates a species closest to *Rhodotorula glutinis*

Next we assembled longer *Rhodotorula* contigs from each SSc sample individually. We used PRICE for the assembly as it is designed to assemble paired-end reads in a complex metagenomic dataset into contigs (Ruby et al. 2013). We seeded contig assembly with *Rhodotorula* reads. The four normal skin samples had insufficient numbers of *Rhodotorula* reads for contig assembly. Each SSc sample generated at least one contig which aligned to *R. glutinis* 28S rRNA. The D1–D2 hypervariable region at the 5′ end of 28S rRNA of *R. glutinis* was covered in each sample. As sequence for this region is available for a wide variety of *Rhodotorula* fungal species, we used this area for further phylogenetic analysis. A multiple sequence alignment was performed using this region of 28S from our four SSc samples as well as selected sequences from NCBI for *Rhodotorula* and related fungal species. This alignment was used to create a phylogentic tree (Figure 4). The general structure of this tree is quite similar to other published phylogenetic trees for Rhodotorula (Biswas et al. 2001). The four sequences from the SSc samples cluster together, with sequences from *R. mucilaginosa*, *R. glutinis*, and *R. graminis*.

In addition, we aligned the original read sets against *R. glutinis* 28S rRNA (NCBI record FJ345357) and viewed the resulting alignment in IGV (Thorvaldsdóttir et al. 2013) (Figure 5). The alignment shows many more *Rhodotorula* reads in the SSc samples. In addition, the reads aligning in the SSc samples have fewer sequence differences from the NCBI record, suggesting their source is more similar to *R. glutinis* than the reads aligning from the normal samples. In addition, the normal samples have six bases in the region from 900–1050 bp that show variability within each normal sample, suggesting multiple *Rhodotorula* species may present.

PRICE was able to assemble additional contigs whose best alignment was to *Rhodotorula* genes, though the longest contig in each sample was a sequence whose best alignment in the nt database was to *R. glutinis* 28S rRNA (FJ345357). Given that rRNA is present at much higher quantities than other transcripts, this is not unsurprising. Other genes assembled were likewise genes expected to be present at high levels, such as *R. glutinis* 18S rRNA (HQ420261.1), *R. mucilaginosa* 18S rRNA (X84326.1) and *R. taiwanensis* RS1 complete mitochondrial genome (HF558455.1) as identified by best hit in the nt database. *R. taiwanensis* is a recently identified, novel *Rhodotorula* species closely related to *R. mucilaginosa* and *R. glutinis var dairenensis* whose entire mitochondrial genome was recently sequenced (Huang et al. 2011) (Zhao et al. 2013). It is one of the few *Rhodotorula* mitochondrial genomes in the database and likely indicative *R. glutinis* mitochondrial sequences, which are absent from the database.

Discussion

Previous studies have suggested a role for pathogens as a trigger of SSc, though neither the pathogen nor the mechanism of pathogenesis is known (Grossman et al. 2011). Here we present RNA-seq data on the microbial species present in skin samples from four patients presenting with early, diffuse SSc and four normal patients. Human reads were filtered from the read sets and the resulting reads were aligned to the NCBI nt database to quantify the non-human species present as inferred by the number of reads aligning to each species. While the quantity of most microbial species showed no difference between normal and SSc samples, *R. glutinis* levels were significantly higher in SSc samples. While normal patients had almost no detectable *R. glutinis* sequences (mean score 0.02 per million reads), patients with SSc had a mean *R. glutinis* score of 5.04 per million reads. Further, we demonstrate that a 28S rRNA sequence most similar to *R. glutinis* can be assembled from each of the SSc samples. Additional studies are needed to definitively determine which species of *Rhodotorula* are present in these samples, although at the level of sequence available from our RNA-seq assemblies, the species appears to be most similar to *R. glutinis* (NCBI record FJ345357) though there appears to also be similarity to *R. mucilaginosa*. This is not surprising since these two species are closely related.

Rhodotorula are environmental yeast found in soil, air, lake and seawater as well as peanuts, fruit juices, crustaceans and mollusks (reviewed in (Wirth and Goldani 2012)). *Rhodotorula* can also be found on plastic shower curtains, toothbrushes, humidifiers and dishwashers (Alvarez-Fernández et al. 1998) (Zalar et al. 2011). *Rhodotorula* are an opportunistic pathogen, particularly as a cause of central venous catheter and peritoneal dialysis-associated fungemia (Tuon and Costa 2008). Disseminated fungemia can occur in immunocompetent and immunocompromised hosts (Tuon and Costa 2008). *Rhodotorula* has also been reported in localized infection of the skin and lung of humans and animals (Alvarez-Fernández et al. 1998), (Kayman et al. 2013) (Monga and Garg 1980). *R. glutinis* cell wall preparations can stimulate macrophage activation *in vitro*, suggesting that this yeast might drive pulmonary inflammation (Sorenson et al. 1998), and hypersensitivity pneumonitis has been reported with inhaled *Rhodotorula* (Alvarez-Fernández et al. 1998). A recent rat model of disseminated *R. mucilaginosa* infection revealed involvement of the lungs, liver and spleen with a granulomatous inflammatory reaction (Wirth and Goldani 2012). *Rhodotorula*-associated peritoneal fibrosis has been reported in patients with dialysis-associated fungemia (Eisenberg et al. 1983), suggesting that inflammation-driven fibrosis in the skin or lung is a potential consequence of infection by this fungal species.

These data also raise the hypothesis that *Rhodotorula* colonization is a consequence of skin disease, rather than a trigger of systemic sclerosis. This cross-sectional study will inform future longitudinal studies of the mycobiome of patients with systemic sclerosis. Latrogenic immunosuppression may also predispose patients to fungal infections regardless of the underlying disease; however two of the SSc patients in this study had not been treated with immunosuppressive agents. Future studies may examine the mycobiome in other fibrotic skin diseases and in latrogenic immunodeficiency.

All biopsies in this study were taken from the forearm. We used control biopsies from healthy patients rather than clinically unaffected skin from SSc patients as previous studies have demonstrated molecular changes in the unaffected skin of SSc (Whitfield et al. 2003; Milano et al. 2008). Recent studies have shown that the mycobiome of the skin varies by site (Findley et al. 2013); future research in this area will require a survey of lesional and clinically unaffected skin from a variety of body sites. Future studies will also be needed to determine which layer of the skin is colonized by Rhodotorula.

One possible explanation of these results is lab contamination in the RNA-seq, however we believe this is unlikely. The four SSc samples were collected on different dates. In at least one case, a normal was collected at the same time, under the same conditions as the SSc sample (the normal was the spouse of the SSc patient). Samples were treated identically from the point of collection onwards. It is difficult to imagine a scenario where the SSc samples could have become contaminated in the lab without the normal samples being similarly affected.

Future research will be required to test this hypothesis and fulfill Hill's epidemiologic criteria for causal association. It is crucial to demonstrate that pathogen exposure precedes development of SSc, which will require prospective studies. The modern genomics view of Koch's postulates stipulates that fewer copies of pathogen nucleic acid exist in normal tissue, consistent with our data. Longitudinal studies will be needed to demonstrate that pathogen load correlates with disease severity and resolution or relapse. Finally, molecular and cellular research is needed to determine how this pathogen triggers inflammation and fibrosis in SSc. Preclinical animal models will allow *in vivo* research on the effect of pathogen on disease.

Materials & Methods

Sample collection

All study participants gave written, informed consent under a Boston University Medical Center Institutional Review Board approved protocol. The study conformed to the Declaration of Helsinki Principals. Single 4 mm punch biopsies were obtained from lesional forearm skin of four patients with early, diffuse SSc within 6 months of first onset of non-Raynaud's symptoms, and normal forearm skin of four controls without disease in a design similar to our original microarray studies (Whitfield et al. 2003). Tissue was stored in RNAlater at −80° C until processed. Samples were processed at Dartmouth Geisel School of Medicine under a protocol approved by the Committee for the Protection of Human Subjects (CPHS) at the Geisel School of Medicine.

RNA-seq sample preparation and sequence alignment

Total RNA was extracted from skin biopsies using QIAgen RNeasy Plus Mini kit. A modified protocol was used to isolate both large (>200 nt) and small (<200 nt) RNA fraction. Only the large RNA fraction was used for this study. Ribosomal RNA was depleted from the large RNA fraction using Invitrogen Ribominus kit. RNA-Seq library was synthesized by NuGen Ovation RNA-Seq System v2 using cDNA prepared by random

hexamer priming. Libraries were multiplexed and sequenced on an Illumina HiSeq 2000 platform and 187–242 million 50 bp paired-end reads were obtained per sample.

Metagenomic analysis was performed with the Integrated Metagenomic Sequence Analysis (IMSA) package (Dimon et al. 2013). The samples were also analyzed by DNA microarray and assigned to the intrinsic gene expression subset (Milano et al. 2008). All four of the SSc patients were assigned to the inflammatory intrinsic subset defined by correlation to centroids. RNA-seq data from these eight patients are available at NCBI GEO at accession number GSEXXXXX (*In process, number will be added in proof*).

Filtering low quality and human reads

Human reads were filtered from the RNA-seq read sets using IMSA (Dimon et al. 2013). Reads were quality filtered to remove any reads with more than 3 bases with a quality score below 15. Next, human reads were removed by progressively more stringent alignments to the human genome (hg19), first with bowtie, then with blat, followed by blast. The final blast alignment removed all reads aligning to the human genome with an E-value of 1xe-8 or better.

Initial Taxonomic Analysis

Once the human reads had been filtered from the dataset, the reads were aligned to NCBI's nt database using blast (E-value <= 1e-15). Scores were generated at every level of NCBI's taxonomy to classify the resulting sequences. Reads aligning to a single classification would add a score of one to that classification. For reads that aligned to multiple species with equal scores, each species was given a partial score (i.e. if a read aligned to two species equally, each would get a score of 0.5 from the read; if the read aligned to three species equally, each would receive 0.33 from the read).

For further analysis, only reads with a single best alignment were retained to avoid spurious hits. In addition, scores were normalized to the millions of reads in the initial read set. Cluster analysis was done using only bacterial, fungal and viral genera with a score above 0.01 in at least one sample. Unsupervised clustering was performed using Cluster 3.0 (Eisen et al. 1998) and visualized using TreeView (Saldanha 2004).

The TaxMap bubble diagrams were created by IMSA, visualized using GraphViz (Gansner and North 2000). TaxMaps show the score at each taxonomic level, in this case for the unique reads only. An average TaxMap was created for Normal and SSc samples by considering only the nodes present in all four samples of the given type, calculating the mean value across the four samples, then only displaying nodes with a score above 0.05 to make the graphs more readable. Individual unfiltered TaxMaps for each sample can be found in the supplemental info for both Plant/Fungal reads and for Bacterial reads.

To align reads to *R. glutinis* 28S rRNA (NCBI FJ345357), we used Bowtie (Langmead et al. 2009) to align the full RNA-seq dataset against the NCBI record. Results were converted to a sorted BAM file and visualized in IGV (Thorvaldsdóttir et al. 2013).

Assembled sequences

For each SSc sample, longer *Rhodotorula* sequences were assembled using PRICE (Ruby et al. 2013). Normal samples did not contain enough *Rhodotorula* reads for assembly. The assembly was run for each sample individually, seeded with all the reads aligning uniquely to *Rhodotorula* (NCBI tax id 5533) (specific flags: -nc 40 -mol 40 -tol 20 -mpi 90 -target 90 1 1 1).

To determine the source of contigs assembled by PRICE, a BLAST alignment to NCBI's nt database was performed using the online BLAST website. Every SSc sample contained a sequence aligning to the 5′ end of the 28S transcript of *Rhodotorula glutinis*. In specific, every sequence covered bases 786–1147 of NCBI record FJ345357 (Khot et al. 2009), spanning the D1–D2 hypervariable region of the 28S rRNA. Multiple sequence alignment was performed in MUSCLE (Edgar 2004). A phylogenetic tree was created using Phylogeny.fr, which uses MUSCLE for the sequence alignment then constructs a maximum likelihood tree using PhyML (Dereeper et al. 2008).

Acknowledgments

This work was supported by grants from the Scleroderma Research Foundation (SRF) to MLW and NIH P50 AR060780 to RL and MLW. ZL received support from NIH R01 AR061384 (MLW) and DoD CDMRP PR100338P1 (MLW). TW also received support from NIH P30 AR061271 to RL and MLW. MEJ also received support from a SYNERGY pilot grant from Geisel School of Medicine at Dartmouth.

Abbreviations

SSc	systemic sclerosis, scleroderma
IMSA	Integrated Metagenomic Sequence Analysis

References

Alvarez-Fernández JA, Quirce S, Calleja JL, et al. Hypersensitivity pneumonitis due to an ultrasonic humidifier. Allergy. 1998; 53:210–2. [PubMed: 9534923]

Biswas SK, Yokoyama K, Nishimura K, et al. Molecular phylogenetics of the genus Rhodotorula and related basidiomycetous yeasts inferred from the mitochondrial cytochrome b gene. Int J Syst Evol Microbiol. 2001; 51:1191–9. [PubMed: 11411687]

Chung L, Fiorentino DF, Benbarak MJ, et al. Molecular framework for response to imatinib mesylate in systemic sclerosis. Arthritis Rheum. 2009; 60:584–91. [PubMed: 19180499]

Dereeper A, Guignon V, Blanc G, et al. Phylogeny.fr: robust phylogenetic analysis for the nonspecialist. Nucleic Acids Res. 2008; 36:W465–469. [PubMed: 18424797]

Dimon MT, Wood HM, Rabbitts PH, et al. IMSA: integrated metagenomic sequence analysis for identification of exogenous reads in a host genomic background. PLoS ONE. 2013; 8:e64546. [PubMed: 23717627]

Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research. 2004; 32:1792–1797. [PubMed: 15034147]

Eisen MB, Spellman PT, Brown PO, et al. Cluster analysis and display of genome-wide expression patterns. PNAS. 1998; 95:14863–8. [PubMed: 9843981]

Eisenberg ES, Alpert BE, Weiss RA, et al. Rhodotorula rubra peritonitis in patients undergoing continuous ambulatory peritoneal dialysis. Am J Med. 1983; 75:349–52. [PubMed: 6881189]

Findley K, Oh J, Yang J, et al. Topographic diversity of fungal and bacterial communities in human skin. Nature. 2013; 498:367–70. [PubMed: 23698366]

Gansner ER, North SC. An open graph visualization system and its applications to software engineering. SOFTWARE - PRACTICE AND EXPERIENCE. 2000; 30:1203–33.

Greenblatt MB, Sargent JL, Farina G, et al. Interspecies Comparison of Human and Murine Scleroderma Reveals IL-13 and CCL2 as Disease Subset-Specific Targets. Am J Pathol. 2012

Grossman C, Dovrish Z, Shoenfeld Y, et al. Do infections facilitate the emergence of systemic sclerosis? Autoimmun Rev. 2011; 10:244–7. [PubMed: 20863912]

Hamamdzic D, Kasman LM, LeRoy EC. The role of infectious agents in the pathogenesis of systemic sclerosis. Curr Opin Rheumatol. 2002; 14:694–8. [PubMed: 12410093]

Hinchcliff ME, Huang CC, Wood TA, et al. Molecular Signatures in Skin Associated with Clinical Improvement During Mycophenolate Treatment in Systemic Sclerosis. J Invest Dermatol. 2013 In Press.

Huang C-H, Lee F-L, Tien C-J, et al. Rhodotorula taiwanensis sp. nov., a novel yeast species from a plant in Taiwan. Antonie Van Leeuwenhoek. 2011; 99:297–302. [PubMed: 20680683]

Kayman T, Sarigüzel FM, Koç AN, et al. Etiological agents of superficial mycoses in Kayseri, Turkey. J Eur Acad Dermatol Venereol. 2013; 27:842–5. [PubMed: 22672104]

Khot PD, Ko DL, Fredricks DN. Sequencing and analysis of fungal rRNA operons for development of broad-range fungal PCR assays. Appl Environ Microbiol. 2009; 75:1559–65. [PubMed: 19139223]

Langmead B, Trapnell C, Pop M, et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009; 10:R25. [PubMed: 19261174]

Milano A, Pendergrass SA, Sargent JL, et al. Molecular subsets in the gene expression signatures of scleroderma skin. PLoS ONE. 2008; 3:e2696. [PubMed: 18648520]

Monga DP, Garg DN. Ovine pulmonary infection caused by Rhodotorula rubra. Mykosen. 1980; 23:208–11. [PubMed: 7402216]

Ouellette BF, Boguski MS. Database divisions and homology search files: a guide for the perplexed. Genome Res. 1997; 7:952–5. [PubMed: 9331365]

Pendergrass SA, Lemaire R, Francis IP, et al. Intrinsic gene expression subsets of diffuse cutaneous systemic sclerosis are stable in serial skin biopsies. J Invest Dermatol. 2012; 132:1363–73. [PubMed: 22318389]

Ruby JG, Bellare P, Derisi JL. PRICE: software for the targeted assembly of components of (Meta) genomic sequence data. G3 (Bethesda). 2013; 3:865–80. [PubMed: 23550143]

Saldanha AJ. Java Treeview—extensible visualization of microarray data. Bioinformatics. 2004; 20:3246–8. [PubMed: 15180930]

Sargent JL, Milano A, Bhattacharyya S, et al. A TGFbeta-responsive gene signature is associated with a subset of diffuse scleroderma with increased disease severity. J Invest Dermatol. 2009; 130:694–705. [PubMed: 19812599]

Sorenson WG, Shahan TA, Simpson J. Cell wall preparations from environmental yeasts: effect on alveolar macrophage function in vitro. Ann Agric Environ Med. 1998; 5:65–71. [PubMed: 9852493]

Steen VD, Medsger TA. Changes in causes of death in systemic sclerosis, 1972–2002. Ann Rheum Dis. 2007; 66:940–4. [PubMed: 17329309]

Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013; 14:178–92. [PubMed: 22517427]

Tuon FF, Costa SF. Rhodotorula infection. A systematic review of 128 cases from literature. Rev Iberoam Micol. 2008; 25:135–40. [PubMed: 18785780]

Whitfield ML, Finlay DR, Murray JI, et al. Systemic and cell type-specific gene expression patterns in scleroderma skin. Proc Natl Acad Sci USA. 2003; 100:12319–24. [PubMed: 14530402]

Wirth F, Goldani LZ. Experimental Rhodotorulosis infection in rats. APMIS. 2012; 120:231–5. [PubMed: 22339681]

Zalar P, Novak M, de Hoog GS, et al. Dishwashers--a man-made ecological niche accommodating human opportunistic fungal pathogens. Fungal Biol. 2011; 115:997–1007. [PubMed: 21944212]

Zhao XQ, Aizawa T, Schneider J, et al. Complete mitochondrial genome of the aluminum-tolerant fungus Rhodotorula taiwanensis RS1 and comparative analysis of Basidiomycota mitochondrial genomes. Microbiologyopen. 2013; 2:308–17. [PubMed: 23427135]

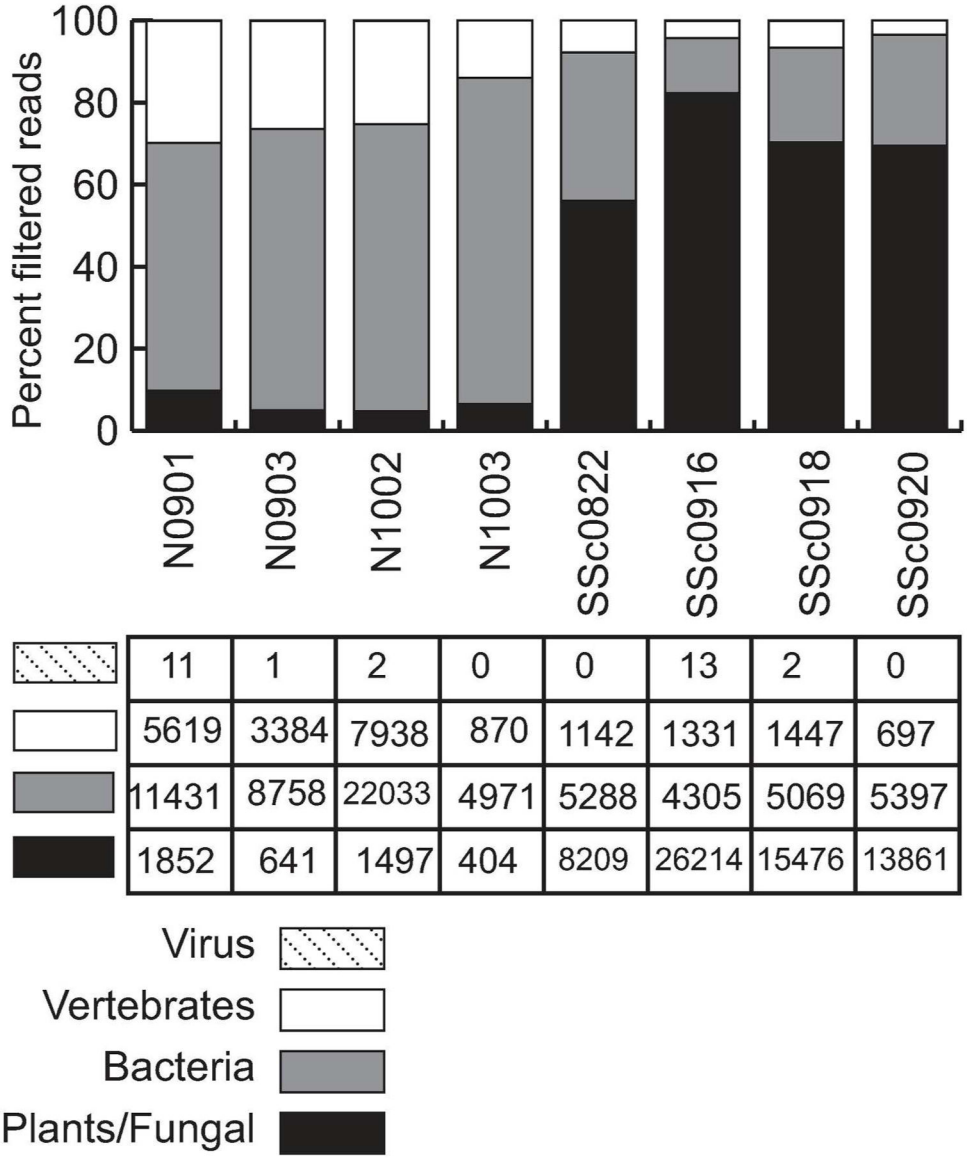

Figure 1. IMSA analysis reveals plant/fungal sequences in SSc samples

Division-level breakdown of the reads remaining after filtering human reads. NCBI divisions group plants and fungal sequences together. Normal samples begin with “N” while SSc samples begin with “SSc”. Numbers shown in the table below are the IMSA score for each division per million input reads, indicating the relative abundance as a proportion of total reads. The most striking difference between normal and SSc samples is the abundance of plant/fungal reads in SSc samples, with an associated reduction in other divisions. Vertebrate reads (primarily human reads not filtered by IMSA due to mismatches to the human genome) are about a quarter of the non-plant/fungal reads. The other three-quarters of the non-plant/fungal reads are bacteria, the amount of which varies by sample but does not have a clear difference in abundance between normal and SSc samples.

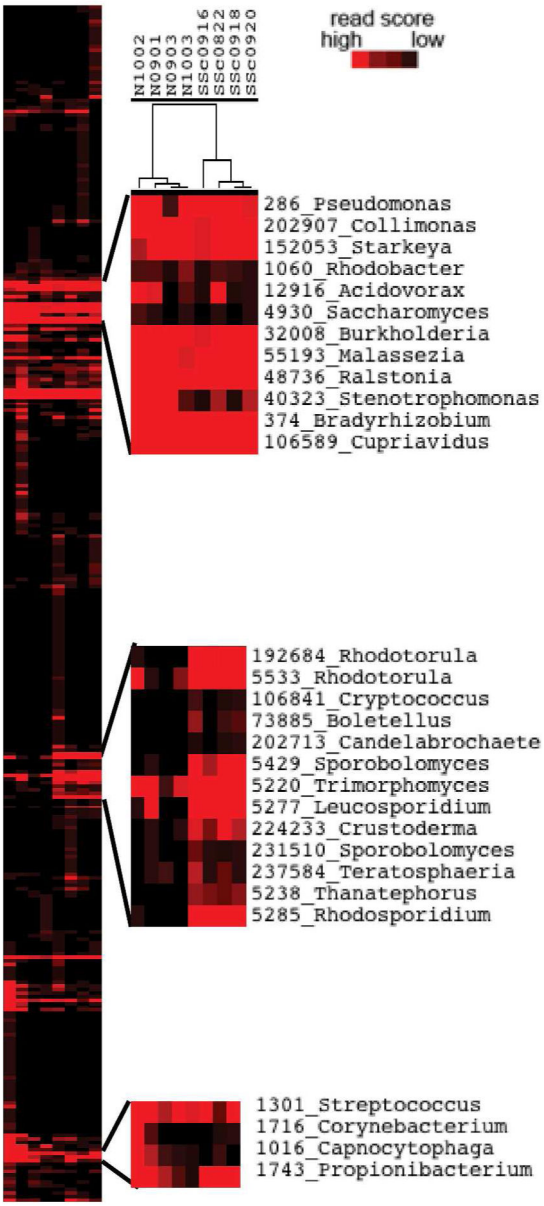

Figure 2. Unsupervised clustering of genera scores

All bacterial, fungal and virual genera with a score above 0.01 in at least 1 sample were clustered by IMSA score, normalized per million reads in the original read set. The tree shows normal samples cluster together on the left while SSc samples cluster together on the right. The top and bottom call-outs show normal skin flora expressed in both normal and SSc samples. The center call-out shows fungal genera expressed predominantly in SSc samples.

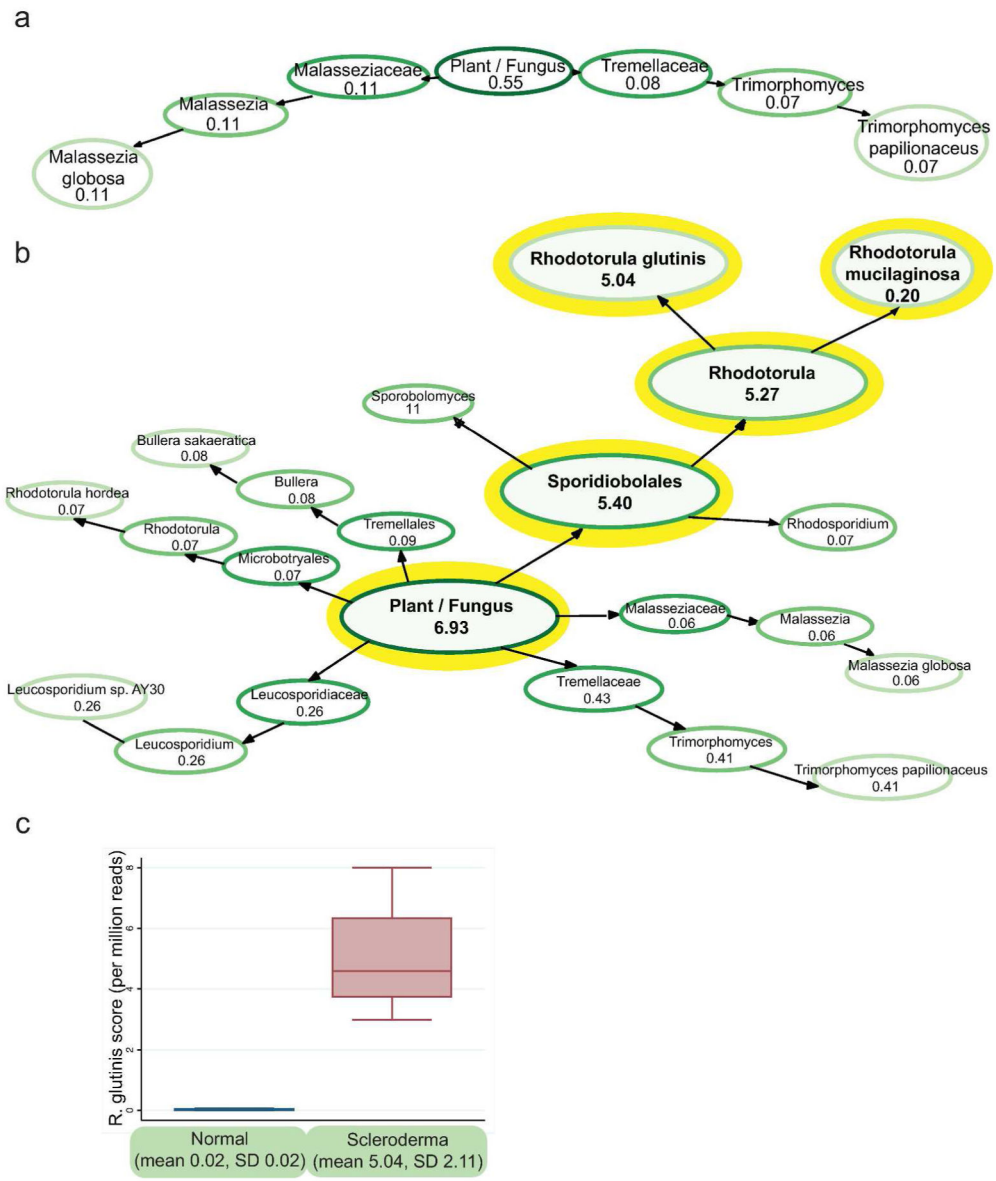

Figure 3. *Rhodotorula glutinis* in Normal and SSc skin samples

a. TaxMap visualization of plant/fungal reads shows few fungal species in normal skin. TaxMaps show the IMSA score for each level of the taxonomic hierarchy, allowing quick visualization of the metagenome of a sample. The score shown is the average score for the normal samples counting only reads with a single best alignment, normalized per million input reads. Only nodes with a value above 0.05 are shown to make the figure easier to read. b. TaxMap visualization demonstrates *R. glutinis* as the dominant species in SSc skin. c. Reads with a single best alignment to *R glutinis* are present at 252-fold higher frequency in SSc skin than normal skin.

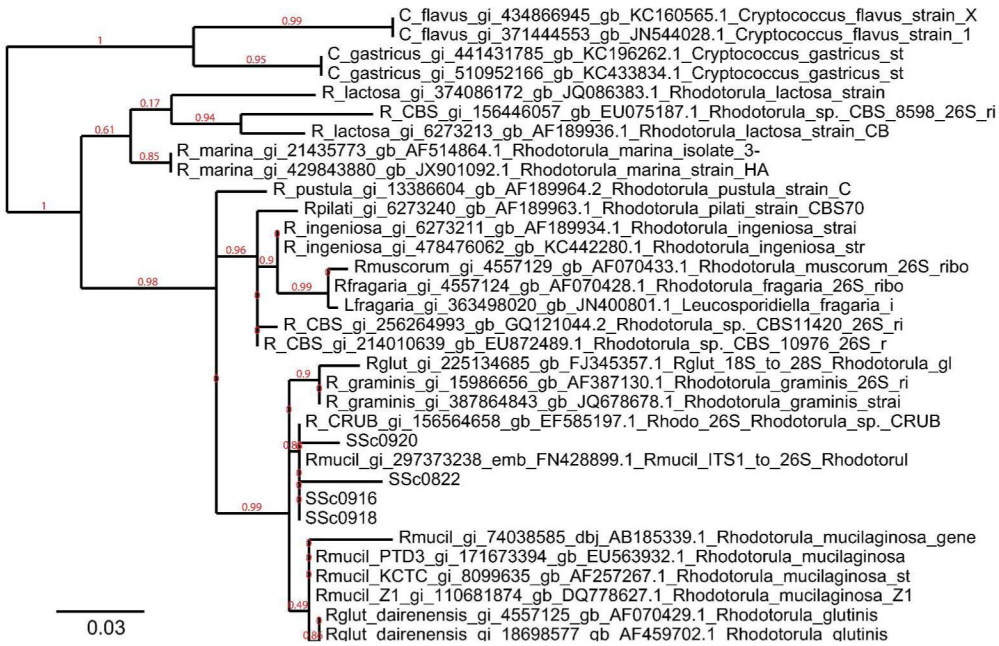

Figure 4. Phylogenetic tree of assembled 28S rRNA sequences

Phylogenetic tree of 28S rRNA sequences from selected NCBI *Rhodotorula* sequences. The maximum likelihood tree was constructed with PhyML and rendered with TreeDyn. Species names and accession numbers are given for sequences downloaded from GenBank. The four SSc samples are grouped with *R. mucilaginosa*, close to sequences from *R. glutinis* and *R. graminis*.

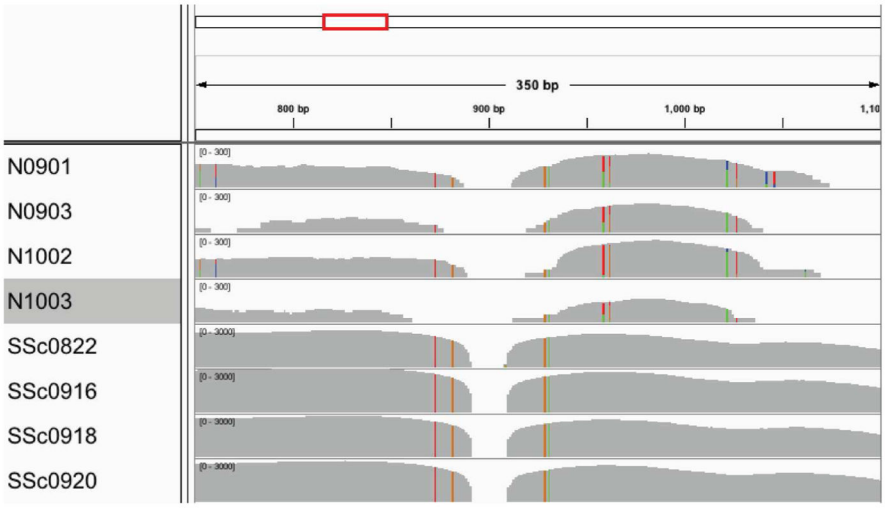

Figure 5. IGV visualization of original read set to *R. glutinis* 28S rRNA

Aligning raw reads to *R. glutinis* 28S rRNA sequence (FJ345357) shows many reads aligning in the SSc samples but much fewer reads in the normal samples. The position shown is from 750–1100 in the sequence, which is the end of ITS2 and the first 400 bases of 28S rRNA. The gray histogram shows the depth of coverage at each base along the sequence; note that the axis is 10-fold higher for SSc samples. Colored bars show areas where the aligned reads differ from the reference sequence.

Table 1

Clinical features of SSc patients

Healthy controls included two white males and two white females, ages 24, 27, and unknown.

	sex	race	age	diagnosis	disease duration (months)	MRSS	immunosuppression
SSc0882	M	W	60	dcSSc	7	38	none
SSc0916	F	W	56	dcSSc	36	44	none
SSc0918	M	W	50	dcSSc	3	15	prednisone
SSc0920	F	W	65	dcSSc	6	34	mycophenolate

MRSS: modified Rodnan skin score, M: Male, F: Female, W: White, dcSSc: diffuse cutaneous systemic sclerosis.

Table 2

Sequence read counts and IMSA scores

Rhodotorula scores are normalized per million reads in the initial readset.

Sample	Initial Fastq	High Quality Reads	Reads After Human Filter	*Rhodotorula* Score (normalized)	*R. glutinis* score (normalized)
N0901	188,871,690	157,580,342	498,774	0.005	0.005
N0903	241,690,526	186,982,854	391,620	0.000	0.000
N1002	199,655,946	170,008,890	935,514	0.075	0.065
N1003	192,679,262	110,673,772	222,532	0.016	0.016
SSc0882	203,708,052	146,919,248	304,502	3.102	2.985
SSc0916	228,994,824	155,180,738	301,846	8.253	8.000
SSc0918	243,159,154	171,379,530	337,658	4.684	4.495
SSc0920	188,011,530	126,597,588	256,664	4.888	4.675