Differential metatranscriptomic profiling during *Candida* infections

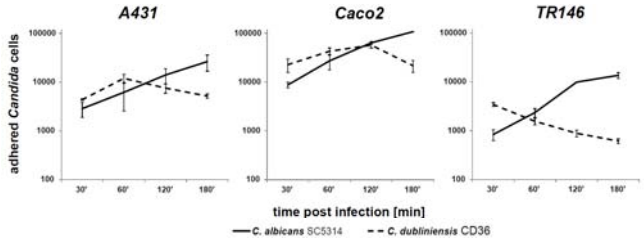
C. Grumaz[1], H. Hoffmann[2], S. Lorenz[2], P. Stevens[2], E. Lindemann[2], S. Rupp[2], K. Sohn[2]

[1] University of Stuttgart IGVT, Nobelstr. 12, 70569 Stuttgart, Germany, [2] Fraunhofer IGB, Nobelstr. 12, 70569 Stuttgart, Germany

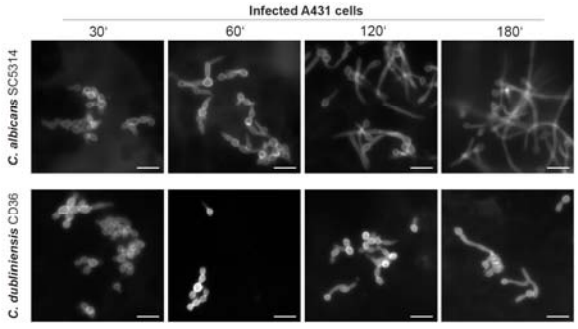
With the advent of next-generation DNA deep sequencing technologies unbiased approaches for the analysis of gene expression including RNA-seq became available. In contrast to DNA-microarrays, the determination of quantitative data using RNA-seq is accomplished by counting individual transcripts rather than by deducing transcript abundances from signal intensities following hybridization to specific probes. This concept with its unique open and unbiased architecture allows not only for experimental *de novo* annotations of whole transcriptomes but also provides most reliable and comprehensive analyses of complex samples derived from more than one species, so called metatranscriptomes, even from unknown organisms. Furthermore, gene expression profiling across closely related species became possible by directly comparing the normalized and semi-absolute quantities of previously defined orthologs. We made use of these advantages and applied this method to the analysis of host-pathogen interactions during infections of human epithelia by *Candida albicans* or *Candida dubliniensis* to simultaneously study host and pathogen responses on the transcriptional level. Homologous as both species are with respect to phylogeny, the more distinct they behave regarding colonization efficiency, morphogenesis and virulence. Accordingly, we investigated the interaction of each species with vulvo-vaginal reconstituted human epithelia (RHE) during early stages of adhesion where *C. albicans* is showing significantly higher adhesion rates than *C. dubliniensis*. For this purpose, complex RNA from the RHE infected with either *C. albicans* or *C. dubliniensis* was isolated after 1h and after 3h of adhesion and subsequently prepared for RNA-seq on Illumina HiSeq2000 with 50bp read length. In average, about 80% of all reads could be uniquely mapped to each species providing quantitative data for each gene following species-dependent normalization. Analyses of the host showed a similar response to both species at the transcription level. To directly compare the profiles across *C. albicans* and *C. dubliniensis*, we also determined the orthologs' quantities and tested for differences in the transcriptional responses resulting in critical differences during adhesion process in both species.

Adhesion rate of *Candida dubliniensis* gradually decreases

In a first step, we established stable RHE (= reconstituted human epithelia) models with three cell lines for infection with *Candida* sp. representing natural habitats in human host – vulvo-vaginal (A431), colorectal (Caco2) and buccal (TR146) epithelia. To analyse early stages of infection with *C. albicans* (SC5314) and *C. dubliniensis* (CD36), we determined their adhesion rates on RHEs. For this, we infected each RHE with 100,000 *Candida* cells and stained the adherent cells with Calcofluor white after certain time points (30', 60', 120' and 180'). Adherent cells were counted microscopically.
The first observation we made is the unequal adhesion efficiency across all cell lines (Caco2>A431>TR146) in both species. The second and thouroughly more striking observation occurs in the adhesion behaviour of *C. dubliniensis* – after a cell line dependent increase in adhesion, a significant decrease follows on every cell line (see lower figure).



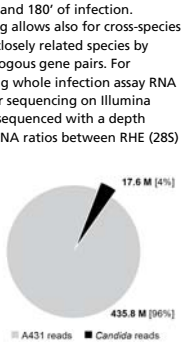
Microscopic images over the analysed time course of both species on the vulvo-vaginal RHE reveals the most obvious reason for the different behaviour in these early stages of adhesion – *C. albicans*' morphology differs from that of *C. dublinienis*. While *C. albicans* forms true hyphae between 60' and 120', *C. dubliniensis* primarily remains as blastospore with sporadic pseudohyphae only after 180' (see lower figure). These morphology differences were observed on Caco2 and TR146 RHEs as well.

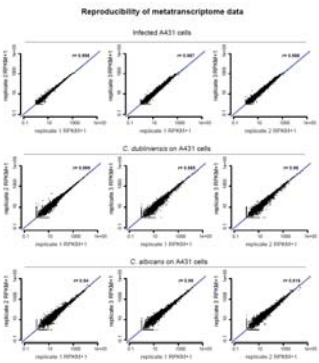


However, it is not yet clear why *C. albicans* forms hyphae while *C. dubliniensis* remains in the blastospore morphology under identical infection conditions, although both are phylogenetically closely related. One possibility might be that the respective genetic repertoires are significantly different or that conserved genes are regulated in a different manner. The other possibility could be an active and discriminative response of the human epithelia according to the respective species. We tried to check these hypotheses by analyzing their interactions on the transcriptional levels.

Deep sequencing of metatranscriptomes derived from infected RHEs

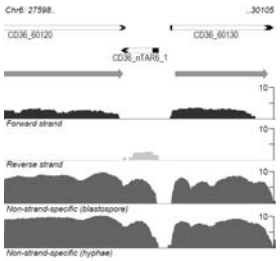
We applied deep sequencing to the analysis of the above-named host-pathogen interactions between vulvo-vaginal RHE and *C. albicans* or *C. dubliniensis* to simultaneously study host and pathogen responses on the transcriptional level after 60' and 180' of infection.
Furthermore, deep sequencing allows also for cross-species gene expression profiling of closely related species by previously determined orthologous gene pairs. For metatranscriptome sequencing whole infection assay RNA was isolated and prepared for sequencing on Illumina HiSeq2000. Each sample was sequenced with a depth corresponding to the large rRNA ratios between RHE (28S) and *Candida* (25S).
In general, RHE RNA was massively overrepresented and about 40-50 M reads per sample were needed to result in 1-2 M reads for *Candida* quantification. Reads were then mapped against the respective genome databases (see right figures).

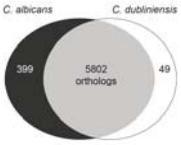




Reevaluation of the *in silico* predicted gene models and the defined orthologs across *C. albicans* and *C. dubliniensis*

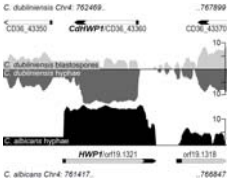
In a previous study, we reevaluated the *in silico* predicted gene models by collecting experimental data by using two complementary sequencing technologies – FLX for long and strand-specific reads along with Illumina for short reads in high amounts. Following genome-associated assembly of sequence reads we were able to generate experimentally verified databases containing 6016 and 5972 genes for *C. albicans* and *C. dubliniensis*, respectively. Comparison of our databases with publically available gene models for *C. albicans* and *C. dubliniensis* confirmed approximately 95% of already predicted ORFs, but also revealed so far unknown novel transcriptionally active regions (nTARs) in both species, coding and non-coding (example is shown in right figure: visualization of sequence reads and annotation in a genome browser).



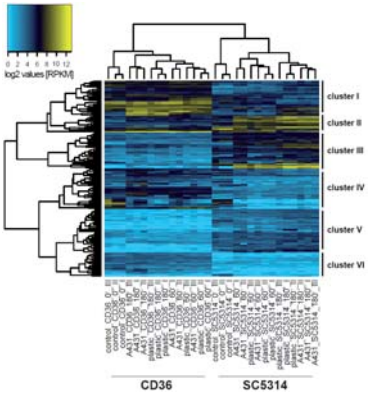


Qualitative cross-species analysis of these databases revealed in addition to 5802 coding orthologs also 399 and 49 species-specific coding genes for *C. albicans* and *C. dubliniensis*, respectively (see left figure). Additionally, comparison of ncRNAs resulted in 131 non-coding, conserved orthologous pairs across both species. 318 and 331 nc-nTARs seem to be specific for *C. alibicans* or for *C. dubliniensis*, repectively (not shown).

Using deep sequencing as an unbiased and open technology for gene expression profiling, it enables the possibility to directly compare transcript abundances of orthologs across two or more species under identical conditions. An example for the orthologous pair of *HWP1*/orf19.1321 (=hyphal wall protein 1) is illustrated in the right figure visualized in a genome browser and showing high expression rates during hyphal growth in both species.



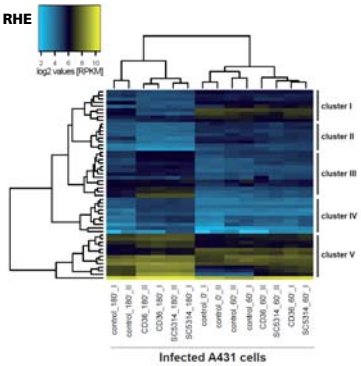
Differential expression of orthologs (DEO) in *C. albicans* and *C. dubliniensis*



Following mapping, quantification and cross-species normalization of transcript numbers, differential gene expression was calculated. Significant differential gene expression was found for 173 genes with a foldchange > $|\log_2 3|$ in at least one comparison across the species (0' control, 60' on A431 cells, 60' on plastic, 180' on A431 cells or 180' on plastic).
The heatmap to the left is illustrating (i) that *C. dubliniensis* infections are clustered separately from *C. albicans* and (ii) that 60' and 180' conditions cluster independently from substrate – RHE or plastic.
Moreover, we found defined clusters with higher expression rates for *C. dubliniensis* (cluster I, IV & VI) as well as for *C. albicans* (cluster II, III &V).

Differential expression in human vulvo-vaginal RHE

Downstream analyses for RHE data was performed like for *Candida* data and checked for differential gene expression.
Significant differential gene expression was found for 53 genes with a foldchange > $|\log_2 1|$ in at least one comparison of the different infection timepoints to the 0' control. In summary, the transcriptional reaction of the epithelia seem to be quite weak during these early stages of infection. However, the heatmap to the right shows no discriminative response between the two species. While cluster I and II indicate groups of genes being similarly downregulated, cluster III-V show comparably upregulated genes in response to *C. albicans* or *C. dubliniensis* infection.



contact
Christian Grumaz / Kai Sohn
Phone: +49 711 970 4055
Christian.Grumaz@igb.fraunhofer.de
Kai.Sohn@igb.fraunhofer.de

Fraunhofer IGB
Nobelstr. 12,
70569 Stuttgart, Germany
www.igb.fraunhofer.de

University of Stuttgart
Institute for Interfacial Engineering

Fraunhofer IGB