Identifying the common interaction networks of amoeboid motility and cancer cell metastasis

Ahmed H. Zeitoun1, Shady S. Ibrahim2, Christoph P. Bagowski3
1German University Cairo, Faculty of Pharmacy and Biotechnology, New Cairo City, Egypt
2ETH Zürich, Institute of Biochemistry, Schafmattstrasse 18, 8093 Zürich, Switzerland
3Ernst-Moritz Arndt University Greifswald, Friedrich Löffler Straße 23a, 17475 Greifswald, Germany
E-mail: bagowskic@uni-greifswald.de

Received 6 February 2012; Accepted 10 March 2012; Published online 1 June 2012

Abstract
The recently analyzed genome of Naegleria gruberi, a free-living amoeboflagellate of the Heterolobosea clade, revealed a remarkably complex ancestral eukaryote with a rich repertoire of cytoskeletal-, motility- and signaling-genes. This protist, which diverged from other eukaryotic lineages over a billion years ago, possesses the ability for both amoeboid and flagellar motility. In a phylogenomic comparison of two free living eukaryotes with large proteomic datasets of three metastatic tumour entities (malignant melanoma, breast- and prostate-carcinoma), we find common proteins with potential importance for cell motility and cancer cell metastasis. To identify the underlying signaling modules, we constructed for each tumour type a protein-protein interaction network including these common proteins. The connectivity within this interactome revealed specific interactions and pathways which constitute prospective points of intervention for novel anti-metastatic tumour therapies.

Keywords movement; invasion; protein interaction network; PDZ domain; Proteome; cancer; tumour; interactome; prostate; breast carcinoma; metastatic behaviour.

1 Introduction
Naegleria gruberi is a heterotrophic protist living primarily in its amoeba form in both freshwater and wet soils around the globe (De Jonckheere, 2002). Naegleria gruberi is best known for its remarkably fast differentiation from amoebae to flagellates with two anterior 9+2 flagella (Fulton, 1993). A transformation, which includes the construction of a complete new cytoplasmic microtubule cytoskeleton (Fulton, 1993).

In addition, this amoeboflagellate can further exist as a resting cyst, which ultimately will produce the amoebae form (Fulton, 1993) and it possess several key features that distinguish eukaryotes from prokaryotes including a complete actin and microtubule cytoskeleton and an extensive transcriptional machinery (Fulton and Walsh, 1980).

The genomes of several other free-living protists have been sequenced (e.g., Dictyostelium (Eichinger and Noegel, 2005), Thalassiosira (Armbrust et al., 2004), Tetrahymena (Eisen et al., 2006), Paramecium (Aury et al., 2006), Chlamydomonas (Merchant et al., 2007), however the discovery of the Naegleria gruberi genome as a distant clade, allows for substantially broader phylogenomic comparisons of free-living eukaryotes than possible before (Fritz-Laylin et al., 2010).
Micromonas, a member of the picophytoplankton group, is a free-living eukaryote found throughout marine ecosystems - from polar waters to the tropics (Worden, 2006). The Micromonas strain RCC299 (or NOUM17) (also known as Micromonas pusilla RCC299 or Micromonas sp. RC299) was isolated in 1998 in the Equatorial Pacific and its complete genome was sequenced 11 years later (Worden et al., 2009). Important in the context of this study is, that Micromonas RCC299 owes its motility to one single flagellum.

Flagellar motility exists in all major eukaryotic groups (including the Heterolobosea clade of Naegleria gruberi) and therefore can be considered as an ancestral feature (Cavalier-Smith, 2002). Likewise, actin-based amoeboid locomotion is found in many diverse eukaryotic lineages and this form of motility also originated early in eukaryotic evolution, perhaps even being present in an eukaryotic ancestor (Fritz-Laylin et al., 2010; Cavalier-Smith, 2002; Fulton, 1970).

Fig. 1 (A) Schematic view of the subtraction approach for the identification of amoeboid motility genes in cancer. For three different cancer types (breast carcinoma, prostate tumour and malignant melanoma) common genes with Naegleria gruberi were identified and subtracted were genes present in Micromonas RCC299. Abbreviations used are AM = amoeboid motility and FM = Flagellar motility. Naegleria gruberi is capable of both AM and FM whereas Micromonas only for FM. Our approach selects for genes important for AM found in the different tumour types. (B) Variation of Coverage and Ident parameters and the distribution of matches between Naegleria gruberi and malignant melanoma. Different combinations of parameters for coverage, identities and gaps were tested in order to recognize a comfortable tipping point which may minimize false positives and false negatives. Subsequently, a requirement for 30% coverage, 25% identities and 10% gaps was used in our analyses as parameters for matches.
The amoeboid-like crawling of cells is a highly regulated cellular process which plays crucial roles for migrating cells during development as well as for metastasizing human tumour cells. This amoeboid movement is connected to the actin cytoskeleton dynamics and can be broken down into three essential steps (Fig. 1): protrusion of the leading edge, adhesion to the substrate, and retraction of the cell body (Mitchison and Cramer, 1996). Protrusion involves localized actin polymerization at the leading edge membrane, which ultimately leads to the extension of a lamellipod, a broad sheet-like structure. As the lamellipod advances, adhesions create molecular grips to the substrate. Retraction is the process wherein the adhesions are broken and the cell body is pulled forward. Protrusion is the best understood of these processes (Pollard and Borisy, 2003).

It is noteworthy to mention that amoeboid motility can exist without actin, an example for this exception is found in the nematode sperm where the cytoskeleton is based on the 14-kD major sperm protein, MSP, rather than on actin (Stewart et al., 1998).

A better understanding of the molecular basis of cell motility is central for the identification of crucial mechanisms important for metastatic behaviour of tumour cells. Without question, invasion, migration and metastatic behaviour of cancer cells depends heavily on cell shape changes and cell motility.

The fact that metastasis formation and the arising secondary tumours are responsible for more than 90% of all cancer deaths further stresses this importance (Jemal et al., 2007). In a complex process, primary tumour cells invade neighbouring tissues, intravasate the systemic circulation, transfer through the vasculature and the lymphatics, arrest in distant capillaries and lymphatic vessels, then extravasate into the perivascular tissue, and ultimately proliferate from micrometastases into macroscopic secondary tumours (Fidler, 2003; Marques et al., 2009).

Tumour progression involves a multitude of regulatory genes and signaling modules, of which many are still elusive (Ibrahim et al., 2011; Tacutu et al. 2011). Important during these processes are cell shape and surface changes, the interactions between tumour cells and their microenvironment including their neighbouring cells and of special importance is the motility of the malignant tumour cells (Marques et al., 2009).

Cancer cells that crawl through a 3D matrix can be morphologically distinguished by two modes of invasion: one that appears as a mesenchymal cell movement that relies on proteolytic degradation of the surrounding matrix and another that adopts a rounded, more amoeboid mode of motility that frequently is accompanied by cell blebbing (Wolf et al., 2003; Sahai, 2005) showed that cancer cells invading through 3D collagen under conditions in which matrix degradation is blocked efficiently use an amoeboid mode of invasion with the formation of bleblike constriction rings. Consistent with this, some cancer cells preferentially use the bleb-associated amoeboid mode of invasion to bypass the requirement for extracellular matrix proteolysis or simply to switch modes of motility as an escape mechanism (Sahai and Marshall, 2003; Friedl, 2004).

Furthermore, experimental induction of PM blebs in noninvasive cells promotes their ability to invade into 3D matrices (Gadea et al., 2007; Tournaviti, 2007). One mode of cellular regulation occurs directly at the level of Rho, where activation of Rho causes PM blebbing and overexpression of RhoC induces a primitive amoeboid-like highly invasive phenotype characterized by the formation of dynamic membrane protrusions and blebs (Sander et al., 1999; Gadea et al., 2007).

Although the role of amoeboid motility in cancer invasion and metastasis formation has been recognized, not much is known about the molecular mechanisms important for cell crawling (amoeboid movement) and bleb formation. The goal of this study is to identify the molecular modules and pathways important for amoeboid motility in human tumour cells (a schematic overview of our approach is shown in Fig.1 A).
2 Material and Methods

2.1 Data collection and processing
Proteomic comparisons were done using BLAST (Altschul et al., 1997). To interpret the results, a parser application was developed using AutoIt v3 (ver. 3.3.6.1, http://www.autoitscript.com/) which extracts and reprocesses outputs from any of NCBI’s BLAST web interface, BLAST+ (ver. 2.2.24) and NetBLAST (ver. 2.2.24) into tab-delimited text files which can be imported and manipulated using a variety of statistical packages, such as Microsoft Excel. Unlike software packages developed with the same purpose in mind (e.g. Batch Blast Extractor (Pirooznia et al., 2008), Noblast, Jamblast (Lagnel et al., 2009), our script allowed side-by-side comparisons of multiple BLAST outputs of different organisms/proteomes against the same central proteome of interest. While the script was designed to extract all BLAST output data, only a vital subset (query length, identities % and gaps %) is carried forward into the output file in order to minimize processing load and clutter. In addition to these variables, a new value, “Coverage %” was calculated based on the relative sizes of the query and target sequences in order to allow a better interpretation of the results. Threshold values for Identities, Gaps and Coverage were determined and the highest scoring (E-score) was processed.

2.2 Species comparison and proteome sources
The proteomic data for prostate cancer was derived from five different patients with metastatic prostate carcinoma (Khan et al., 2010). For malignant melanoma a list of over 4000 proteins based on the transgenic mouse tumour model, TG3, was recently described (Zanivan et al., 2008). The melanoma list of mouse proteins was BLASTed against the UniProt human protein database to generate the equivalent human proteins (no additional restrictions were used). Two studies describing the proteomic signatures of metastatic breast carcinoma were combined and used here for comparison (Johann et al., 2009; Wang et al., 2010). All of these different proteomic sources were used for our cross-species comparison to the proteomes of Naegleria gruberi (Fritz-Laylin et al., 2010) and Micromonas RCC299 (Worden et al., 2009).

2.3 Constructing and visualizing the human interactome and the three cancer networks
The HPRD (8th release) list of human binary protein-protein interactions in tab delimited format was obtained from the HPRD website (http://www.hprd.org/). This list was loaded into the Cytoscape software (ver 2.6.3, http://www.cytoscape.org/) to visualize the human interactome network. The three lists of proteins (generated individually for each tumour type e.g. proteins found in the breast carcinoma proteome and in Naegleria gruberi but absent in Micromonas RCC299) were generated as described above. The Uniprot accession numbers and Entrez IDs of these lists were converted into the HPRD ID list format by using the converter at the Biomedical Information Research Center (BIRC) of the Japanese National Institute of Industrial Sciences and Technology (AIST) (http://biodb.jp/ids). After the conversion, the HPRD datasets were used to construct cancer networks on 3 different levels (zero, 1st and second degree networks). The three different zero degree networks (the network including only genes from the respective lists) are shown in Fig. 2. Only the direct interactions between the listed genes were included in this network. The next level cancer networks included the interactions of the respective cancer AM genes with their 1st neighbours, but excluded neighbour-neighbour interactions. The third network was set to include the cancer nodes as well as their 1st and 2nd neighbours. The edges in this network represent the interactions between the cancer genes themselves, between them and their 1st or 2nd neighbours and between 1st and 2nd neighbours, but not the interactions among 2nd neighbours. Due to the large size of the higher level networks and their similarity to each other and the general interactome, further analysis was restricted to the level of the 2nd neighbour and below. Supplementary material shows the connectivity of all constructed networks, and of the entire interactome (Statistics section).
3 Results and Discussion

3.1 Phylogenomic comparison of Naegleria gruberi and Micromonas RCC299

We developed a parser application, which extracts and reprocesses outputs from any BLAST web interface (see Supplementary material tutorial). Our script allowed side-by-side comparisons of multiple BLAST outputs of different organisms, and we used this to compare the genomes of two distinct protists and to further perform a phylogenomic substraction analysis.

Initially, we selected genes which were absent in Micromonas RCC299 and present in Naegleria gruberi. The former a monoflagellate, only able to move through its single flagellum and the latter an amoeboflagellate with the ability for both types of locomotion (amoeboid motility (AM) as well as flagellar motility (FM)). In this first step, we anticipated to enrich for amoeboid-motility-associated genes.

Obviously, any phylogenetic analysis generated in this manner, will exclude genes that are used both for motility and other processes e.g., alpha-tubulin, which is used in flagella but also mitotic spindles (Fritz-Laylin et al., 2010) and further actin, Arp2/3 (which nucleates actin filaments), or other general actin cytoskeletal components, as these genes are found across eukaryotes regardless of their capacity for amoeboid locomotion (Fritz-Laylin et al., 2010). Nonetheless, some of these form the basis for locomotion, through their involvement in the dynamic cytoskeletal reorganization underlying amoeboid motility. Additionally, it is expectable that our two way phylogenomic comparison will have many false positives (genes not involved in amoeboid motility). Considering these aspects, we have incorporated two additional means for filtering in this study, which are des cribed and discussed below.

The proteomes utilized, retrieved from the National Center for Biotechnology Information website in July, 2010, were comprised of 20280 proteins and 31581 proteins for each of Micromonas RCC299 and Naegleria gruberi, respectively.

3.2 Comparing different human tumour entities to the gene set enriched for amoeboid motility

Subsequently, we use the set of genes found present in Naegleria gruberi but absent in Micromonas RCC299, to compare them directly to the “proteomic” profiles of three different tumour entities. The proteomic data for prostate cancer was derived from five different patients with metastatic prostate carcinoma. The experimental strategy coupled isobaric tags for relative and absolute quantitation with multidimensional liquid phase peptide fractionation followed by tandem mass spectrometry. Over 1000 proteins were quantified across the specimens (Khan et al., 2010).

For proteomic profiling of malignant melanoma, the transgenic mouse tumour model, TG3 was employed. These mutant mice carry skin melanomas and under defined experimental conditions, a total of 100 μg of solid tumour lysate yielded a melanoma proteome of 4443 identified proteins (Zanivan et al., 2008).

For generating the proteomic signatures of metastatic breast carcinoma, human lymph node containing breast carcinoma metastasis of six specimens (~ 50,000 cell each) were collected by Laser Capture Microdissection and analysed by biological mass spectrometry. 367 proteins were identified in this approach (Johann et al., 2009). This proteome was further supplemented by data acquired from the biological spectrometric analysis of MCF-7 circulating tumor cells (Wang et al., 2010).

Out of the 4437 malignant melanoma proteins we find here 340 present in Naegleria gruberi but absent in Micromonas RCC299, likewise out of 1031 prostate proteins investigated 105 were found and of the 1050 breast carcinoma proteins tested 93 were found present in Naegleria gruberi but absent in Micromonas RCC299 (Supplementary material Statistics section). Found common among all three tumor entities were 29 proteins (Table 1) (28 of these 29 proteins had HPRD IDs with the two exceptions: ACAT1 and LMAN2, which therefore do not appear in the common network).
Fig. 2 Interactions linked to amoeboid motility (AM) of human cancer types. Shown are the cancer networks for breast carcinoma (A), for malignant melanoma (B) and for prostate cancer (C). Shown are the 0 degree networks of the three tumour types. Indicated are self loops and in green are identified proteins linked to AM without direct interactions to other AM proteins. Highest connectivity is indicated in red.
Fig. 3 General networks based on the sets of AM linked proteins identified for the different cancer entities. Shown are the cancer networks for breast carcinoma (A), for malignant melanoma (B) and for prostate cancer (C). Shown are the 1 degree networks of the three tumour types including their first neighbours from the general interactome.
Fig. 4 Interaction and pathways identified in breast carcinoma, malignant melanoma and prostate cancer. Shown are examples of interactions identified in breast carcinoma (A), for malignant melanoma (B) and for prostate cancer (C). Identical interactions were found for breast and prostate cancer.
Fig. 5 Common identified proteins and their interactions. Shown are the 29 genes and their interactions identified in all three tumour types and in \textit{Naegleria gruberi} but not found in \textit{Micromonas RCC299}. 
Table 1 Twenty nine proteins of three tumor entities

<table>
<thead>
<tr>
<th>Accession</th>
<th>Gene name</th>
<th>Synonyms</th>
<th>Approved name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q13085</td>
<td>ACACA</td>
<td>ACC1</td>
<td>acetyl-CoA carboxylase alpha</td>
</tr>
<tr>
<td>Q96FG8</td>
<td>ACAT1</td>
<td>THIL</td>
<td>acetyl-CoA acetyltransferase 1</td>
</tr>
<tr>
<td>Q96BG6</td>
<td>ACTN4</td>
<td>FSG1</td>
<td>actinin, alpha 4</td>
</tr>
<tr>
<td>Q59FV6</td>
<td>ACTR3</td>
<td>ARP3</td>
<td>ARP3 actin-related protein 3 homolog</td>
</tr>
<tr>
<td>O43681</td>
<td>ASNA1</td>
<td>ARSA-I, GET3, TRC40</td>
<td>arsA arsenite transporter, ATP-binding, homolog 1 (bacterial)</td>
</tr>
<tr>
<td>P52907</td>
<td>CAPZA1</td>
<td></td>
<td>capping protein (actin filament) muscle Z-line, alpha 1</td>
</tr>
<tr>
<td>P47756</td>
<td>CAPZB</td>
<td></td>
<td>capping protein (actin filament) muscle Z-line, beta</td>
</tr>
<tr>
<td>Q9P110</td>
<td>CPNE3</td>
<td></td>
<td>copine III</td>
</tr>
<tr>
<td>Q7Z4W1</td>
<td>DCXR</td>
<td>DCR, KIDCR, SDR20C1,</td>
<td>dicarboxyl/L-xylulose reductase</td>
</tr>
<tr>
<td>Q53XC0</td>
<td>EIF2S1</td>
<td>EIF-2alpha, EIF2A</td>
<td>eukaryotic translation initiation factor 2, subunit 1 alpha</td>
</tr>
<tr>
<td>P38117</td>
<td>ETFB</td>
<td>FP585</td>
<td>electron-transfer-flavoprotein, beta polypeptide</td>
</tr>
<tr>
<td>P00367</td>
<td>GLUD1</td>
<td>GDH</td>
<td>glutamate dehydrogenase 1</td>
</tr>
<tr>
<td>Q5TZX1</td>
<td>GNA13</td>
<td>87U6</td>
<td>guanine nucleotide binding protein (G protein), alpha inhibiting activity polypeptide 3</td>
</tr>
<tr>
<td>Q71UM7</td>
<td>GNB1</td>
<td></td>
<td>guanine nucleotide binding protein (G protein), beta polypeptide 1</td>
</tr>
<tr>
<td>P06396</td>
<td>GSN</td>
<td>DKFZp313L0718</td>
<td>gelosin</td>
</tr>
<tr>
<td>P55084</td>
<td>HADHB</td>
<td>MTPB, MSTRP029</td>
<td>hydroxyacyl-CoA dehydrogenase/3-ketoacyl-CoA thiolase/enoyl-CoA hydratase (trifunctional protein), beta subunit</td>
</tr>
<tr>
<td>Q99714</td>
<td>HSD17B10</td>
<td>ABAD, CAMR, ERAB, MHB2, MRPP2, SDR5C1, HADH2</td>
<td>hydroxysteroid (17-beta) dehydrogenase 10</td>
</tr>
<tr>
<td>Q96GW1</td>
<td>HSP90B1</td>
<td>GP96, GRP94</td>
<td>heat shock protein 90kDa beta (Grp94), member 1</td>
</tr>
<tr>
<td>Q75874</td>
<td>IDH1</td>
<td>PICD</td>
<td>isocitrate dehydrogenase 1 (NADP+), soluble</td>
</tr>
<tr>
<td>P48735</td>
<td>IDH2</td>
<td></td>
<td>isocitrate dehydrogenase 2 (NADP+), mitochondrial</td>
</tr>
<tr>
<td>Q12907</td>
<td>LMAN2</td>
<td>GP36B, VIP36</td>
<td>lectin, mannose-binding 2</td>
</tr>
<tr>
<td>Q99497</td>
<td>PARK7</td>
<td>DJ-1, DJ1</td>
<td>parkinson protein 7</td>
</tr>
<tr>
<td>P00558</td>
<td>PGK1</td>
<td>PGKA, MIG10</td>
<td>phosphoglycerate kinase 1</td>
</tr>
<tr>
<td>P14618</td>
<td>PKM2</td>
<td>OIP3, PK2, PK3, PKM, THBP1</td>
<td>pyruvate kinase, muscle</td>
</tr>
<tr>
<td>P60900</td>
<td>PSMA6</td>
<td>IOTA, MGC2333, p27K, PROS27</td>
<td>proteasome (prosome, macropain) subunit, alpha type, 6</td>
</tr>
<tr>
<td>Q15185</td>
<td>PTGES3</td>
<td>cPGES, p23, TEBP</td>
<td>prostaglandin E synthase 3</td>
</tr>
<tr>
<td>Q00325</td>
<td>SLC25A3</td>
<td>PHC</td>
<td>solute carrier family 25 (mitochondrial carrier; phosphate carrier), member 3</td>
</tr>
<tr>
<td>Q8WVV9</td>
<td>SRP9</td>
<td></td>
<td>signal recognition particle 9kDa</td>
</tr>
<tr>
<td>O43592</td>
<td>XPO1</td>
<td>XPO3</td>
<td>exportin, tRNA (nuclear export receptor for tRNAs)</td>
</tr>
</tbody>
</table>

29 genes were identified which were found in all three tumour types and in Naegleria gruberi but were absent from Micromonas RCC299. Gene name according to Human Gene Nomenclature (HGNC).

3.3 Creating protein-protein interaction networks for the identified gene sets

Since this comparative analysis will include some false positive genes not linked to amoeboid-motility, we added the involvement in the amoeboid-motility protein-protein interaction network as another criteria. Through this novel approach we identify not single genes, but whole signaling modules and pathways and we anticipate that this method is likely to exclude most of the false negatives and positives of the regular species comparisons. We created three individual networks for prostate cancer, breast carcinoma and malignant melanoma, respectively and one network consisting of the thirty genes common among all three tumour entities and present in Naegleria gruberi but absent from Micromonas RCC299 (Fig. 5).
Acknowledgement
We would like to thank Professor Stephan Verleysdonk for his initial help with the experimental design and input.

References
Eichinger L, Noegel AA. 2005. Comparative genomics of *Dictyostelium discoideum* and *Entamoeba histolytica*. Current Opinion in Microbiology, 8: 606-611
zebrafish xenotransplantation model. BMC Cancer, 9: 128
Worden AZ. 2006. Picoeukaryote diversity in coastal waters of the Pacific Ocean. Aquatic Microbial Ecology, 43: 165-175