Amino Acids (2015) 47:2521–2531
DOI 10.1007/s00726-015-2042-x

ORIGINAL ARTICLE

Composition of distinct sub-proteomes in *Myxococcus xanthus*: metabolic cost and amino acid availability

David E. Whitworth[1] · **Susan E. Slade**[2] · **Adrian Mironas**[1]

Received: 2 April 2015 / Accepted: 29 June 2015 / Published online: 11 July 2015

Abstract Subsets of proteins involved in distinct functional processes are subject to different selective pressures. We investigated whether there is an amino acid composition bias (AACB) inherent in discrete subsets of proteins, and whether we could identify changing patterns of AACB during the life cycle of the social bacterium *Myxococcus xanthus*. We quantitatively characterised the cellular, soluble secreted, and outer membrane vesicle (OMV) sub-proteomes of *M. xanthus,* identifying 315 proteins. The AACB of the cellular proteome differed only slightly from that deduced from the genome, suggesting that genome-inferred proteomes can accurately reflect the AACB of their host. Inferred AA deficiencies arising from prey consumption were exacerbated by the requirements of the 68 %GC genome, whose character thus seems to be selected for directly rather than via the proteome. In our analysis, distinct subsets of the proteome (whether segregated spatially or temporally) exhibited distinct AACB, presumably tailored according to the needs of the organism's lifestyle and nutrient availability. Secreted AAs tend to be of lower cost than those retained in the cell, except for the early developmental A-signal, which is a particularly costly sub-proteome. We propose a model of AA reallocation during the *M. xanthus* life cycle, involving ribophagy during early starvation and sequestration of limiting AAs within cells during development.

Keywords Myxobacteria · Coding bias · Sporulation · Outer membrane vesicles · Secretome · Comparative proteomics

Introduction

Organisms exhibit a wide range of genomic %GC, which strongly influences amino acid composition bias (AACB). It is thought that genomes tend towards high %AT as a consequence of random mutation (Hershberg and Petrov 2010). High %GC is thus thought to be a consequence of selective pressures, although why high %GC is favoured under certain conditions is still not clear (Mann and Chen 2010).

In specific ecosystems, a narrow range of %GC values are observed, resulting from changes in community structure rather than mutation of strains (Foerstner et al. 2005), i.e. selection favours organisms of a particular %GC in distinct environments. Those environments often also promote expression of distinct subsets of genes; so potentially, environment-specific subsets of genes could have different AACB tailored to those environments, with differences in bias likely to be especially profound when comparing environments with different nutritional statuses (Bragg and Wagner 2006; Elser et al. 2011; Merchant and Helmann 2012; Chen et al. 2013).

While it is known that particular genomes of different %GC are selected for/against in different ecosystems, the mechanisms responsible for causation of the selective

Handling Editor: P. R. Jungblut.

Electronic supplementary material The online version of this article (doi:10.1007/s00726-015-2042-x) contains supplementary material, which is available to authorized users.

✉ David E. Whitworth
dew@aber.ac.uk

[1] Institute of Biological, Environmental and Rural Sciences, Aberystwyth University, Ceredigion SY23 3DD, UK

[2] School of Life Sciences, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK

pressure remain obscure (Bohlin 2011). This is mainly due to difficulties in categorising all the variables inherent in an ecosystem, precluding meaningful correlation of particular ecosystem parameters with %GC. For analysis of AACB in subsets of proteins encoded within a microbial genome, the problem is somewhat inverted. While it is possible to control and simplify the environment in the laboratory, there remains difficulty in defining the functional subsets of proteins expressed under distinct environmental conditions.

In this study, we asked whether AACB existed between different subsets of proteins in the model myxobacterium *Myxococcus xanthus* (Whitworth 2008). *M. xanthus* is predatory and its preferred nutrients are AAs (Bretscher and Kaiser 1978). Prey cell lysis is promoted by the secretion of soluble proteins and outer membrane vesicles (OMVs), which are released from *M. xanthus* cells and seem to have roles in both predation and intercellular signalling (Rosenberg et al. 1977; Whitworth 2011). When starved, a population of *M. xanthus* undergoes co-operative development, culminating in the formation of a fruiting body filled with metabolically dormant spores (Velicer and Vos 2009). During the developmental process, an early intercellular signal is A-signal, which is a secreted subset of AA (Kuspa et al. 1992a, b). A-signal appears to be unique in the bacterial world and its nature seems counter-intuitive—using food as a secreted signal of starvation. We were therefore interested in why that particular subset of AA was used as a signal and hypothesised that A-signal composition would be related to AACB usage during different phases of the life cycle.

As part of the study, we undertook whole proteome quantification of three sub-cellular fractions of a *M. xanthus* culture, which allowed us to validate computational results against experimental observations and to investigate AACB in sub-proteomes devoted to nutrient acquisition. The AACB of our experimental dataset closely mirrored that of the deduced proteome, and we found that different sub-proteomes indeed have distinctive AACB. Our results suggest a model of AA reallocation and recycling at different stages of the *M. xanthus* life cycle involving ribophagy during early starvation and sequestration of limiting AA within the cell during development.

Materials and methods

Proteome sample collection

Sample preparation for proteome analysis was based on the method of Evans et al. (2012). Cultures of *M. xanthus* DK1622 were grown at 30 °C in DCY-rich medium (8 mM $MgSO_4$, 10 mM Tris, 2 % casitone, 0.2 % yeast extract, pH 8.0), until they reached late-exponential growth phase. Cells were pelleted by centrifugation (20 min at 10,400*g*), before being washed and resuspended in TM buffer (50 mM Tris, pH 7.8, 10 mM $MgSO_4$). OMVs were removed from the culture supernatant by centrifugation (100,000*g* for 1 h), washed in TM, resedimented and resuspended in TM. Residual OMVs were removed from the culture supernatant by a second centrifugation. OMVs and supernatant preparations (soluble secretome) were filtered (using 0.2 um syringe filters and Amicon ultrafiltration, respectively) prior to storage.

Sample preparation

Samples were then prepared based on the method of Patel et al. (2012) for 2D RP–RP–LC MS^E analysis of tryptic digests, in triplicate. An intact OMV suspension from *M. xanthus* (total volume 20 mL) was heated at 100 °C for 10 min in the presence of 3 mg of Rapigest surfactant (Waters Corporation, Manchester, UK). After cooling on ice, an excess of cold acetone (>10-fold) was added to precipitate proteins and stored at −20 °C overnight. The precipitated protein was collected and re-solubilised in a volume of 500 µL aqueous 0.2 % Rapigest and concentrated using an Amicon Ultra-0.5 Centrifugal Filter Unit with Ultracel-3 kDa membrane at 12,000*g* to a volume of ~50 µL. The solution was repeatedly diluted with 100 mM ammonium bicarbonate to 500 µL and concentrated to reduce the concentration of DKxanthene secondary metabolite. On the final step, the solution was concentrated to a volume of 50 µL.

M. xanthus supernatant and cytosolic proteins were precipitated using a >10-fold excess of ice cold acetone and stored at −20 °C overnight. Proteins were sedimented and washed with excess acetone until no further colour was observed. Precipitated protein was collected and re-solubilised in a volume of 500 µL aqueous 0.2 % Rapigest, heated at 100 °C for 10 min and concentrated using an Amicon Ultra-0.5 Centrifugal Filter Unit with Ultracel-3 kDa membrane at 12,000*g* to a volume of ~50 µL.

Reduction was carried out on each sample with dithiothreitol at 60 °C for 15 min at a final concentration of 9 mM. Alkylation was carried out in the dark with iodoacetamide at ambient temperature for 30 min at a final concentration of 16.7 mM and digested with 1:50 (w/w) sequencing grade trypsin (Promega, Southampton, UK) at 37 °C overnight. RapiGest was hydrolysed by the addition of 2 µL 24 M formic acid, which was then incubated at 37 °C for 20 min. The resulting mixture was vortexed and centrifuged through a 0.22 µm cellulose acetate membrane filter (Corning Incorporated, Corning, NY) to remove particulate matter.

LC–MS

Samples for 2D RP–RP–LC were diluted with an internal standard consisting of a tryptic digest derived from glycogen phosphorylase b (MassPREP Digestion Standard, Waters Corporation) in 20 mM ammonium formate, pH 10.0. Injections of each sample (~600 ng cytosolic protein digest, 70 ng supernatant and 135 ng OMV with internal standard) were loaded and all analyses were conducted in technical triplicate.

For 2D RP–RP–LC, high pH reversed phase–low pH reversed phase nanoscale LC–MS separations of tryptic peptides, tryptic digests were loaded onto the first dimension column, XBridge BEH130 C18 NanoEase Column (300 µm × 50 mm, 5 µm) using a 2D nanoACQUITY system equilibrated with solvent A (20 mM ammonium formate pH 10) at 2 µL/min. The first fraction was eluted from the first-dimension column at 13.1 % B (100 % acetonitrile) for 4 min at a flow rate of 2 µL/min and transferred to the second-dimension trap column via a 1:10 dilution with 99.9 % solvent A (0.1 % formic acid in water) at 20 µL/min, over a time period of 20 min with peptides further resolved on the analytical column. Subsequent fractions were isolated by increasing the concentration of organic solvent from the first dimension. Subsequent first-dimension fractions were eluted at 17.7 and 50.0 % buffer B (0.1 % formic acid in acetonitrile (pH 2.7). Sample loading was increased three- or sixfold dependent on the number of fractions utilised for second-dimension separations.

Precursor ion masses and associated fragment ion spectra of the tryptic peptides were acquired using a Synapt HDMS G2 mass spectrometer (Waters Corporation), using an applied capillary voltage of 3.2–3.5 kV. The time-of-flight analyser of the mass spectrometer was operated in positive ion resolution mode with a typical resolving power of 18,000 calibrated with NaI (1 mg/mL) and CsI (50 ng/mL) or by using the MS/MS spectra of the doubly charged precursor of [Glu1]-Fibrinopeptide B (GFP) (Sigma Aldrich, St. Louis, MO). Data were lockmass corrected using the monoisotopic mass of the doubly charged precursor of GFP sampled from the reference sprayer every 60 s. Accurate mass data were collected in a data-independent mode of acquisition (MSE), whereby trap energy was alternated between 6 V (low) and 10–30 V (elevated). Spectral data were collected over m/z range 50–1950, at an acquisition rate of 0.9, 0.1 s interscan delay. Approximately, 90 % of the precursor and 70 % of the product ions display a mass error of 0 ± 5 ppm.

Protein identification and quantitation

Correlation of precursor ions with their corresponding MS/MS spectra was achieved by the use of a retention time alignment algorithm, followed by a further correlation process based on the physicochemical characteristics of peptides when they undergo CID using ProteinLynx Global Server version 2.5.1 (PLGS-Waters Corporation). Precursor and fragment ion tolerances were determined automatically by PLGS. More than three fragment ions per peptide, seven fragment ions per protein and more than one peptide per protein had to be matched. Protein identifications were accepted if the probability of identification was greater than 95 % and identification of the protein occurred in at least two out of three technical replicate injections.

All data were searched against an *M. xanthus* strain DK 1622 database (http://www.uniprot.org/) containing 7318 sequences. A version of the database was generated by randomizing each protein sequence once, while holding the peptide amino acid composition and the number of tryptic peptides constant, which was then merged with the original database with sequences for internal standards and common contaminants. The false discovery rate was estimated by the number of randomised entries identified (false-positive rate) divided by the number of correct identifications (true-positive rate) expressed as a percentage.

A fixed modification of carbamidomethyl-C was specified. Variable modifications considered included acetyl N-terminus, deamidation N, deamidation Q and oxidation M. Automatic settings for mass accuracy were used and two missed tryptic cleavages were permitted.

Protein quantification was estimated using the Hi3 approach within the PLGS software using signal response from the internal standard glycogen phosphorylase b. An Hi3 absolute quantitation approach is based on the observation that the average MS signal from the three most intense peptides from a protein per mole can be constant (CV ± 10 %). Signal response from an internal standard can be used to calculate a universal response factor (counts/mol) from which the concentration of each protein in each sample be inferred. Each protein concentration (fmol) was initially converted to ng and then calculated as percentage of all the total identified protein concentrations (in ng).

Identified proteins were assigned Enzyme Commission (EC) numbers and Clusters of Orthologous Groups (COG) terms using the IMG database (Markowitz et al. 2012). Proteins were classified as hydrolases if they were assigned an EC number which began with 3, and were denoted as lipoproteins if they were either annotated as lipoproteins, or if LipoP (Rahman et al. 2008) predicted they contained a type II signal peptide. Genetic context was also characterised by dividing genes into operons using the simple criterion that adjacent genes, coded on the same strand of DNA, were considered to be in the same operon. *M. xanthus and M. fulvus* protein sequences were clustered into orthologous groups using the CD-HIT server (Huang et al. 2010) available at http://weizhong-lab.ucsd.edu/cdhit_suite, with a 60 % similarity cutoff.

Proteome AA composition analysis

Deduced proteome sequences (inferred from genome sequences) were extracted via FTP from Refseq for *Bacillus subtilis* str. 168, *M. xanthus* DK1622 and *Escherichia coli* K12. AA frequencies were calculated from sequence data using custom Perl scripts. Proteomes were classified as either quantitative (when the frequency of each protein within the proteome was measured) or qualitative. Qualitative proteomes such as the deduced proteomes were implicitly assumed to be composed of equal quantities of each constituent protein. Quantitative proteomes had the relative abundance of each protein component used as a weighting when calculating AA frequencies.

Spore proteins of *B. subtilis* were defined as those with the word 'spore' in their annotation, while ribosomal proteins of *M. xanthus, B. subtilis and E. coli* were taken as those with 'ribosomal' in their annotation. To calculate the per AA 'metabolic cost' of each proteome, the per AA costs for each AA (as proposed by Akashi and Gojobori 2002) were multiplied by the frequency of each AA in that proteome.

Results

Experimental elucidation of *M. xanthus* sub-proteomes

Samples were taken from a culture of *M. xanthus*, and fractionated into cells, soluble supernatant/secretome and insoluble supernatant/secretome (OMVs), and the proteome of each sample characterised. The composition of the three proteomes is provided in Supplementary File 1 and summarised in Fig. 1. 315 proteins were identified across the three samples (Fig. 1a), with more than three-quarters of those being unique to one of the three fractions. Accumulation curves (Fig. 1b) suggest that any proteome component undetected by our proteomics approach represent at most $\ll 0.1$ % of that proteome. Cellular proteins were enriched relative to the genome-deduced proteome (4.5-fold) in proteins belonging to 'Energy production and conversion', 'Translation, ribosomal structure and biogenesis' and 'Posttranslational modification, protein turnover, chaperones' COG categories. Proteins unique to the cell extract proteome included both GroEL and both EF-Tu paralogues, 22 ribosomal proteins, 5 peptidylprolyl isomerases and numerous enzymes of central metabolism. The secreted proteomes were dominated by proteases, hypothetical proteins and lipoproteins, with a 9.0-fold enrichment of proteins belonging to the 'Amino acid transport and metabolism' COG category, as expected given the preference of *M. xanthus* for amino acids as nutrient sources. Hydrolases were also found abundantly in other studies of *M. xanthus* OMVs, and the proteins identified by us in OMVs overlap significantly with those of Kahnt et al. (2010) and Berleman et al. (2014), although differences in methodology hamper direct comparisons. The secreted proteases belonged to diverse families, implying a broad range of substrate specificity, while the OMV proteome was 3.6-fold enriched for proteins lacking homologues in *M. fulvus*, suggesting there are species-specific differences in predatory activity/mechanisms.

Many proteins were identified in both the OMV and soluble secretomes, and there was a significant correlation between the relative abundances of those proteins in the two fractions ($R = 0.54$). This suggests that much of the soluble secretome could actually be exported from the cell within OMVs which then lyse, rather than through secretion directly into the extracellular milieu. A similarly large overlap between soluble supernatant and OMVs has been observed in *Legionella pneumophila* (Galka et al. 2008).

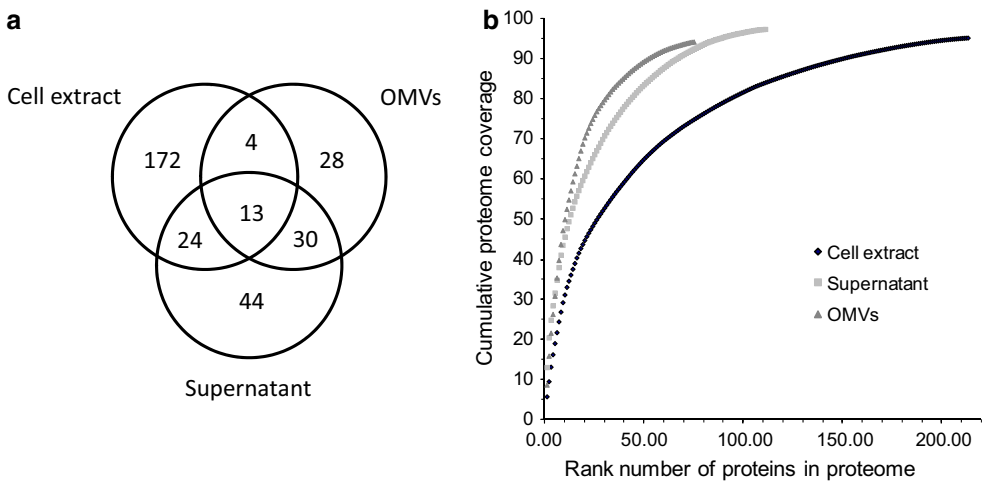

Fig. 1 Experimentally determined sub-proteomes of *M. xanthus*. **a** The number of proteins identified in each of the sub-proteomes and their overlap. **b** Accumulation *curves* showing the % cumulative proteome coverage as a function of the number of proteins identified (ranked by decreasing frequency)

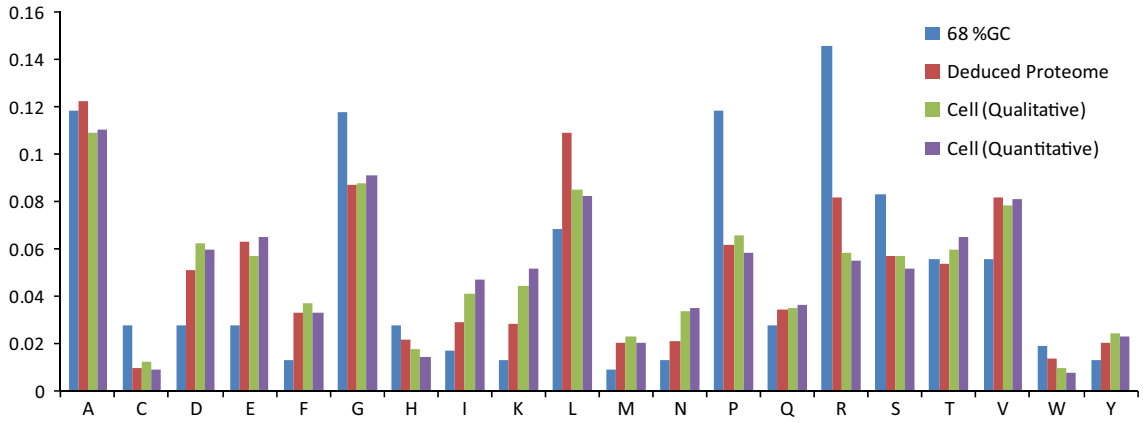

Fig. 2 Relative frequency of AAs encoded by a random 68 %GC genome, the deduced proteome of *M. xanthus* (inferred from the genome sequence) and the cellular proteome of *M. xanthus* (qualitative and quantitative). The deduced proteome is similar to the experimentally characterised proteomes, but very different from that encoded by a 68 %GC genome, being particularly depleted in G, P, R, and S, and enriched in L and V

AA composition of the *Myxococcus xanthus* proteome

The genome of *M. xanthus* is 9.1 Mbp and 68 %GC. The %GC of DNA is known to have a major impact on the AA composition of encoded proteins. We calculated the relative frequencies of AAs expected for a 68 %GC genome, which differed markedly from the AA frequencies observed in the deduced proteome of *M. xanthus* (inferred from the genome sequence). In particular, the deduced proteome was enriched in d, E, f, k, L and V, and depleted in c, G, P, R and S, (Fig. 2), presumably due to selective pressures acting on the different roles of AA within proteins. In this comparison and all subsequent comparisons, AAs are represented by single-letter code in uppercase if the AA frequencies differ by more than 0.025 and lowercase if they differ by at least 0.0125 (corresponding, respectively, to ±50 and 25 % of the average AA frequency). This approach provides an objective approach for comparing frequencies which do not draw disproportionate attention to low-abundance AAs. For ease of interpretation, comparisons described in the text are also shown graphically in Fig. 3.

We also compared the deduced proteome with an experimentally determined proteome—that of the soluble proteins in an extract of *M. xanthus* cells. 'Qualitative AA frequencies' were calculated for the list of proteins identified in the proteome (with every protein assumed to be equally abundant), and each protein's contribution to the AA frequencies was also weighted according to its relative abundance in the proteome to give 'quantitative AA frequencies'. Surprisingly, there was very good agreement between the qualitative and quantitative AA frequencies for the experimentally determined proteome, with no AA frequencies differing by more than 0.007 (Fig. 2). In addition, there was very little difference between the experimental datasets and the deduced proteome—the cellular proteome just being enriched in d, i, k and n, and depleted in a, l and r.

Importantly, these observations suggest that quantification of proteome components does not necessarily affect the apparent AA composition of a proteome, and that 'qualitative' proteomes derived from lists of proteins can be accurate reflections of actual proteome compositions.

The composition of A-signal is distinct from cellular proteomes

A-signal is secreted by a population of starving *M. xanthus* cells, as they begin the process of development which culminates in differentiation of vegetative cells into myxospores. The major spore proteins of *M. xanthus* identified by Dahl et al. (2007) have a very similar AA composition to vegetative cellular proteins (Fig. 4), with just a slight depletion of A residues in the spore (0.014 difference in frequency). Quantitative data are not available for the myxospore proteome; however, the most abundant proteins of the *M. xanthus* spore, proteins C and S (McCleary et al. 1991), show a relative enrichment of K, n, q and depletion of a, h and r relative to the cellular proteome (Fig. 4). Comparing the deduced proteome with spore proteins showed that in *B. subtilis*, it is also true that spores have similar AACB to vegetative cells, with spores enriched in d, h, k and y, and depleted in a, g, i and l.

The AA composition of A-signal as measured by Kuspa et al. (1992a) is very different to both the cellular and spore proteomes. It is enriched in G, i, K, M and S, and depleted in E, f, L, N, P, Q, R, T and v, relative to the cellular proteome (Fig. 4). It is important to note here that the most abundant AAs of A-signal have very little A-signal activity (Kuspa et al. 1992a). Presumably, they act as 'carriers'

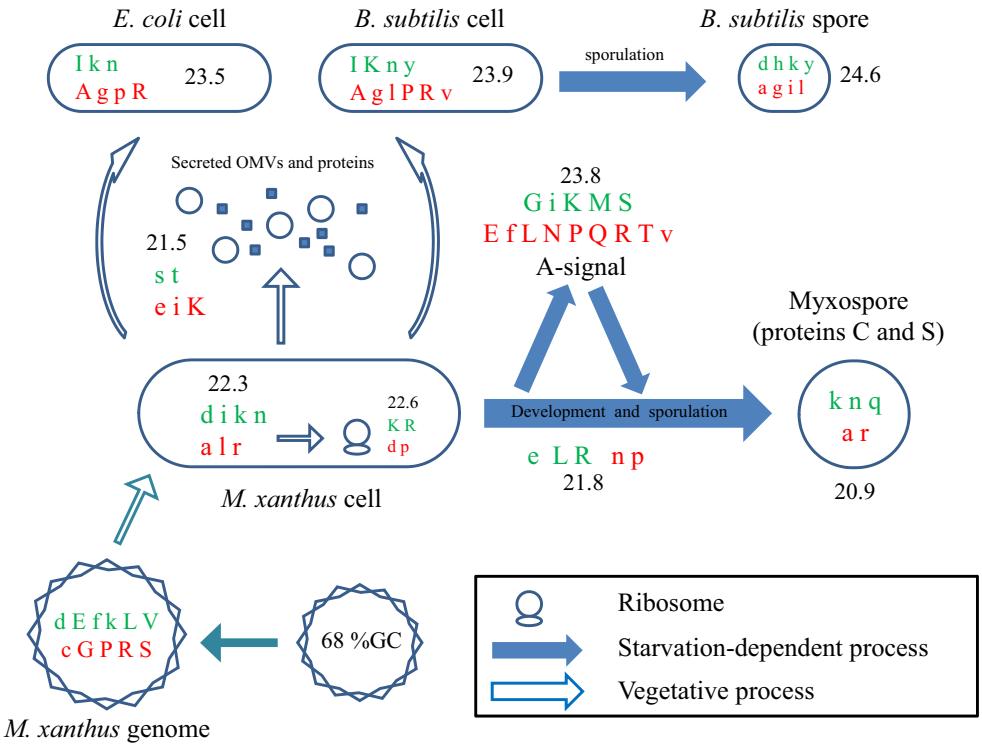

Fig. 3 Differences between the AA composition of sub-proteomes and their relative cost. Each *arrow* represents a comparison between two proteomes and highlights more (*uppercase*) or less (*lowercase*) enriched (*green*) or depleted (red) AA in the arrowhead proteome. For example, the *Bacillus subtilis* spore is enriched in d, h, k and y, and depleted in a, g, i and l relative to the *B. subtilis*-deduced proteome. Most comparisons are made relative to the *M. xanthus*-deduced proteome, but when the *M. xanthus* cell extract proteome is used in the comparison, it is highlighted in the text. *Numbers* refer to the cost of each proteome, as a mean number of ATP per AA

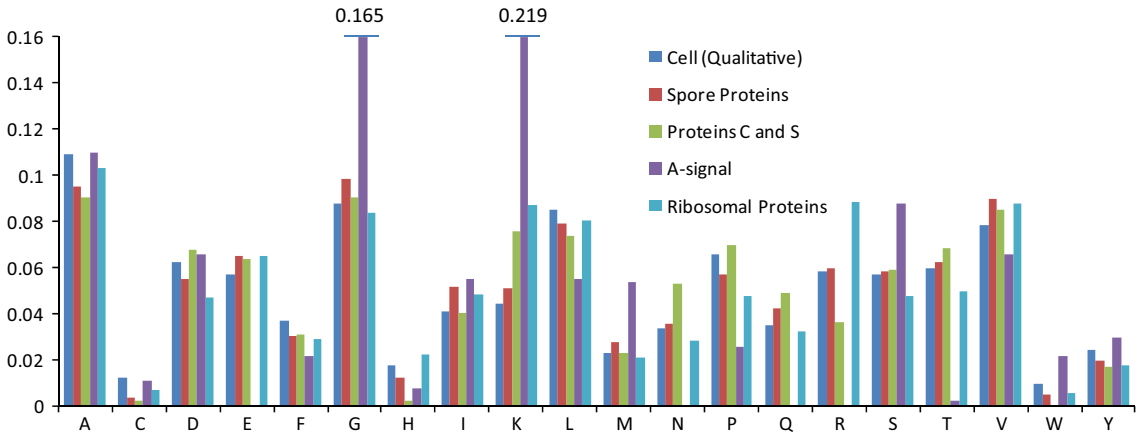

Fig. 4 AA frequencies of the *M. xanthus* cellular proteome (qualitative), spore proteins, A-signal, Proteins C and S and ribosomal proteins. G and K are particularly abundant in the A-signal (off the scale, actual values provided above *bars*). The cellular proteome is similar to that of the spore and major spore proteins, C and S; however, A-signal has a very different composition, being particularly enriched in G, K, M and S. Ribosomal proteins were also found to be enriched in K

for the minor active AA components of A-signal. For the purpose of this topic, it is the abundance of each AA rather than its activity which is important. It is also worth noting that K is highly enriched in A-signal, with a relative abundance/frequency 0.17 higher than in the cellular proteome. There is no obvious relationship between A-signal composition and the spore or cellular proteomes. For instance, K is enriched and N is depleted in A-signal; however, both are enriched in the spore and cellular proteomes.

Ribosomes are particularly rich in the most abundant A-signal AA

A-signal is thought to be made from digested proteins/peptides (Kuspa et al. 1992a, b) and we therefore investigated whether any proteins encoded in the *M. xanthus* genome were particularly enriched for G, K, M and/or S. The 100 proteins with the greatest frequencies of G, M and S (those with 100+ residues) were mostly hypothetical proteins or

lipoproteins (72, 51 and 74 % for G, M and S, respectively, compared to a genome-wide value of 43.6 %). In contrast, the 100 proteins with the greatest %K included just 43 hypothetical proteins and/or lipoproteins, but with 21 ribosomal proteins (the *M. xanthus* genome encodes just 42 proteins annotated as 30S or 50S ribosomal proteins). It is therefore tempting to speculate that A-signal is derived, at least in part, from the digestion of ribosomes, which would be consistent with the shift from vegetative growth to dormancy (Zundel et al. 2009).

The ribosomal proteins of *M. xanthus* are enriched in K and R, and depleted in d and p relative to the cellular proteome (Fig. 4). This is not unique to *M. xanthus*, as the ribosomal proteins of other organisms are also rich in K and R. For instance, *E. coli* ribosomal proteins are enriched in K, R and v (depleted in L), while *B. subtilis* ribosomes are enriched in k and R (and depleted in f). However, there is no direct correlation between ribosomal AA frequencies and A-signal composition, as A-signal is enriched in G and K, and R is absent, whereas ribosomal proteins are enriched in K and R, but not G. Thus, the K of A-signal may derive from ribosome degradation, but A-signal composition must also depend on differential synthesis, degradation, partitioning and/or utilisation of the other AAs.

Secretion and inter-/intracellular signalling

Next, we investigated whether other proteomes partitioned within/outside the cell, or with different internal/external functional roles had different AA compositions. Proteomes of secreted proteins and OMVs were quantified experimentally and compared to that of the cellular proteome. The soluble secreted proteins were enriched for g, n, s and T, and depleted of e, K and l, while OMVs were enriched in s and t, and depleted in e, i and k. Together, the secreted proteome (with an approximately equal contribution from OMVs and soluble secreted proteins, as suggested by our unpublished data) was enriched in s and t, and depleted in e, i and K (Fig. 5). Thus, secreted proteins have a characteristic AACB which differs from that of cellular proteins.

Proteins involved in the regulation of development in *M. xanthus* can be divided into those that regulate internal processes within each cell (intracellular signalling), and those that co-ordinate processes between cells (intercellular signalling). Evans and Whitworth (2010) showed that intracellular and intercellular signalling genes exhibit different properties, so we assessed whether those genes also had different AA compositions. There was very little difference between the two sets of signalling proteins, with the intracellular signalling proteins being enriched in a and l, and depleted in k. However, the combined sets of 'early developmental signalling' proteins when compared to the cellular proteome (qualitative) did have a distinctive AACB, being enriched in e, L and R, and depleted in n and p (Fig. 5).

Predation

Myxococcus xanthus is able to grow in culture when other bacteria or fungi are present as the sole C/N source. Presumably, enough AA could be derived from a prey cell to produce a new *M. xanthus* cell, unless differences in AA frequencies required the interconversion of different AAs for complete incorporation into *M. xanthus* proteins. Comparing the deduced proteomes of two potential prey species with that of *M. xanthus* showed that *E. coli* was enriched in I, k and n, and depleted in A, g, p and R, while *B. subtilis* was enriched in I, K, n and y, and depleted in A, g, l, P,

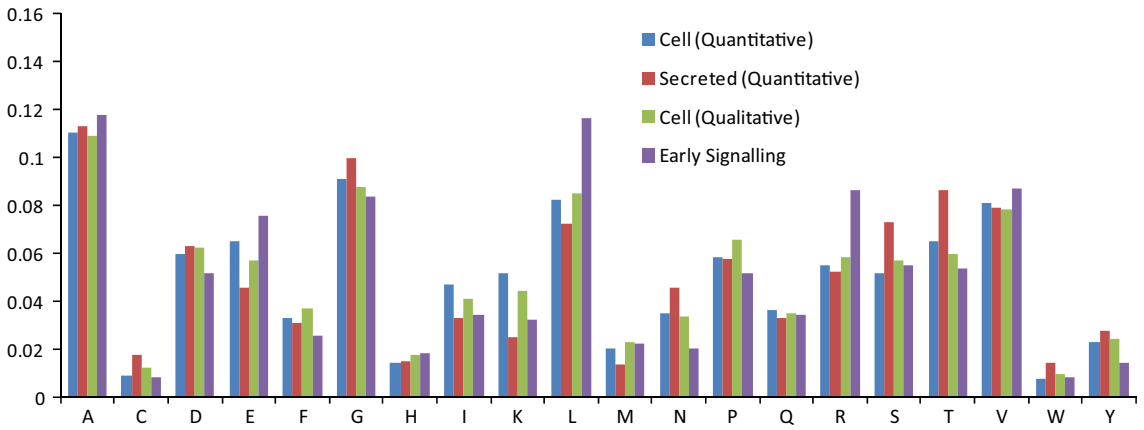

Fig. 5 Relative frequency of AAs in secretomes. Included are the quantitative cellular proteome for comparison with the quantitative secretome, and the qualitative cellular proteome for comparison with the composition of signalling proteins expressed during early development. The secreted proteome mirrors that of the cellular proteome; however early development signalling proteins are enriched in L and R

R and v. Therefore, one might generally expect *M. xanthus* to be limited in A and R while feeding on other bacteria (whether Gram positive or negative), with a relative surfeit of I (Fig. 6). Consistent with this suggestion, the *M. xanthus* cellular proteome is depleted in A and R, but enriched in I relative to the proteome inferred from its genome (Figs. 2, 6).

The metabolic cost of proteomes

While the AAs released by hydrolysis of prey proteins are available for reincorporation into proteins, they could also be metabolised into other AAs or catabolised to provide energy. Different AAs have varying energetic costs for a cell to produce and also yield different amounts of energy when catabolised. Akashi and Gojobori (2002) provide values for the 'cost' of AA in terms of the amount of ATP used/produced by their anabolism/catabolism, and this can be used to calculate an average 'per AA cost' for a proteome (presented in Fig. 3).

The deduced proteomes of *B. subtilis* and *E. coli* are more 'expensive' than that of *M. xanthus* by 1.4 ATP per AA (23.7 compared to 22.3). The secreted proteome of *M. xanthus* is cheaper still, costing an average of 21.5 ATP per AA, and this makes sense from an economic perspective given the loss of those proteins from the cell. Development and sporulation occur in response to nutrient limitation, so it would be expected that spores and developmental proteins would be relatively inexpensive, and this is observed (developmental proteins cost 21.8 ATP per AA and spore proteins cost 21.7 ATP per AA). However, A-signal is noticeably more expensive, at 23.8 ATP per AA, and this seems counter-intuitive given the starving state of the cell and the loss of those AAs by secretion from the cell.

In *B. subtilis,* the spore proteome is actually more expensive than that of the cell. This may reflect the different lifestyles of *B. subtilis* and *M. xanthus,* with germination in *B. subtilis* releasing a single germinant reliant on its own resources to flourish, whereas the *M. xanthus* fruiting body releases a population of germinants able to predate/cannibalise efficiently as a population.

As the typical bacterium is ~55 % protein, a considerable amount of the 10^{10} ATP consumed per cell per hour during growth will be invested in AA metabolism (Feist et al. 2007). Most of the differences in ATP cost per AA that we observe between proteomes appear modest (0.5–1.0). However, with an average of ~22 ATP per AA, even 0.5 ATP per AA difference between proteomes would represent a 2.3 % increase/decrease in efficiency, which will have profound fitness consequences for an organism.

Discussion

In this work, we characterised whole proteomes and subsets thereof in terms of AA frequencies and metabolic cost, and the results of our analyses are summarised graphically in Fig. 3.

In our comparisons, we used a mixture of experimentally determined and computationally generated proteomes. Experimental characterisation of a proteome is non-trivial, and the results are not easy to extrapolate beyond the sample in question. Conversely, genome-derived proteomes are trivial to generate and can be used to investigate biological differences between organisms in a similar way to comparative genomics. The experimentally characterised cellular proteome of *M. xanthus* had a very similar AACB to that of the deduced proteome, whether expressed qualitatively or quantitatively (Fig. 2), suggesting that deduced proteomes

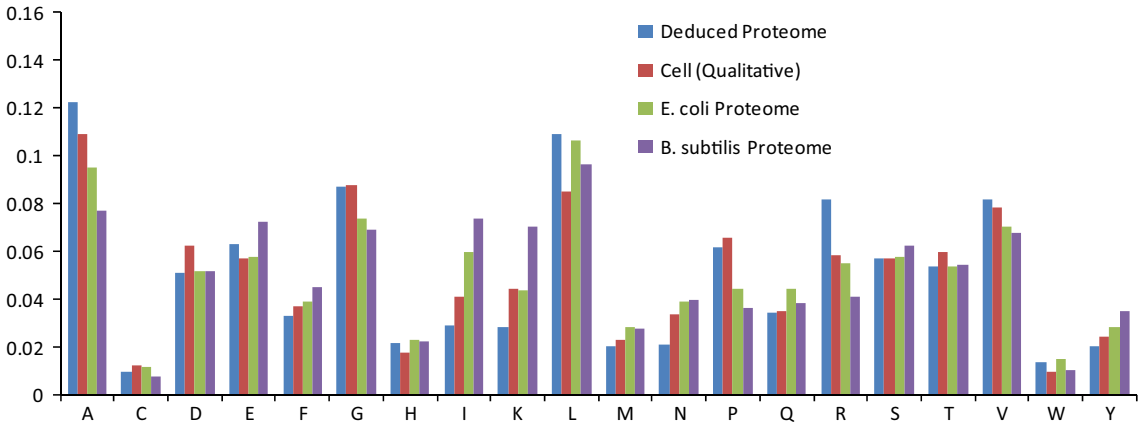

Fig. 6 AA frequencies of nutrient sources. The deduced proteomes of *M. xanthus, E. coli and B. subtilis* are provided alongside the profile for the qualitative cell proteome of *M. xanthus*. Both prey species are relatively depleted in A and R compared to *M. xanthus,* but enriched in I. This is also true of the cellular proteome compared to the deduced proteome (inferred from the genome)

may be good proxies for experimentally elucidated proteomes for studies of AACB.

However, while the qualitative and quantitative cell extract proteomes were very similar in terms of AACB, the similarity masked significant differences between the qualitative and quantitative proteomes. For instance, the COG category 'Coenzyme metabolism', appears to be quite abundant in the qualitative OMV proteome (4 % of the proteome); however in the quantitative analysis, those proteins can be seen to comprise just 0.72 % of the OMV proteome. Similarly, 'DNA replication, recombination and repair' proteins comprise 2.6 % of the qualitative OMV proteome, but just 0.4 % of the quantitative proteome.

The *M. xanthus*-deduced proteome is ' %GC-limited'

The *M. xanthus*-deduced proteome is low in G, P and R which are three of the four AAs encoded by codons beginning with two G/C and which are particularly enriched in high %GC organisms. This implies that the proteome is less '%GC' then the encoding genome, supporting the notion that selection for high %GC occurs directly on nucleotide composition rather than indirectly via AACB (Hershberg and Petrov 2010; Hildebrand et al. 2010). In fact, the *M. xanthus* deduced proteome is 56 '%GC', and it most closely matches that expected of a 56 %GC genome (as judged by a minimum root mean-squared deviation of AA frequencies), while the quantitative proteome most closely matches that predicted for a 52 %GC genome. Deduced proteomes seem to generally have a lower '%GC' than the genome which encodes them, but it is more pronounced for *M. xanthus* than its prey species considered here, which have significantly different %GC (*E. coli* has a 51 %GC genome but a 47 '%GC' proteome, while *B. subtilis* has a 41 '%GC' proteome compared to 43 %GC genome).

The '%GC' of the *M. xanthus* proteome is thus intermediate between that of its prey and that of its own genome. The genome %GC promotes the use of G, R and P; however, those AAs are particularly scarce in its food source.

Recycling/reallocation of AA during the *M. xanthus* life cycle

Patterns of differing AACB led us to propose that ribophagy feeds protein synthesis during early starvation in *M. xanthus*. This proposal is supported by the observation that we identified EF-Tu, GroEL and ribosomal proteins abundantly in vegetative cells, but Dahl et al. (2007) only found EF-Tu and GroELs in spores. During early development, A-signal is secreted, which is rich in I and K, but not R. However, R is noticeably enriched in the cellular proteins expressed during early development, and alongside L. A and P seems to be generally limiting throughout the life cycle, but never appearing to be enriched particularly in sub-proteomes. However, G, the other limiting AA during vegetative growth, is abundantly secreted in A-signal, implying it can be produced in relatively large amounts by the cell (G is one of the three cheapest AAs to produce, according to Akashi and Gojobori (2002)). To generalise, it seems that AAs which are abundant during vegetative growth (K, R, I) are secreted from the cell to produce A-signal during development, while limiting AAs (L and R) are retained within the cell and are enriched in cellular developmental proteins.

Not all AAs can be made/utilised by *M. xanthus* and presumably those AAs must be recycled from prey/autophaged proteins. *M. xanthus* lacks the genes to catabolise I, L and V, which explains its observed *ilv* auxotrophy (Bretscher and Kaiser 1978; Curtis and Shimkets 2008). In addition, it appears to lack full pathways for the catabolism of F, K, W and Y (Curtis and Shimkets 2008), while F, M and N were found to stimulate growth in a defined minimal medium (Bretscher and Kaiser 1978). Several of these limiting/metabolisable AAs are particularly expensive to produce according to Akashi and Gojobori (2002). F, H, I, K, M, W and Y all cost more than 30 ATP per AA to produce, compared to a proteome average of 22.3, four of which (underlined) are auxotrophic or potentially limiting. In addition, F, I, L, W and Y represent five of the six most active A-signalling AAs (Curtis and Shimkets 2008). There thus seems to be a correlation between the ability of *M. xanthus* to metabolise an AA with its metabolic cost, and its activity within the A-signal.

A-signal

The A-signal is a secreted signal required for development, which culminates in sporulation. However, the A-signal's composition does not reflect the general properties of secreted or sporulation proteins. Spore proteins and vegetative secreted proteins are both relatively 'cheap' proteomes, whereas A-signal is notable for its expense. In addition, the abundant AAs of A-signal are not particularly abundant in spore proteins and are depleted in the vegetative secretome and early developmental cellular proteins.

It is tempting to speculate that these properties are important for A-signal activity. The difference in secreted AA between vegetative and starvation conditions means that cells could easily distinguish between nutrient-rich and nutrient-poor conditions by sensing external AA. The A-signal is secreted to a concentration of 0.5 mM (Kuspa et al. 1992a), representing a secretion of 20 million amino acid molecules per cell per hour. With A-signal AA costing 2 more ATP than early developmental proteins, this difference represents investment of an extra ~40 million ATP per cell per hour, which is equivalent to ~2.9 % of the total

cellular maintenance costs of *E. coli* under aerobic conditions (Feist et al. 2007).

Such a significant difference in proteome cost could have profound consequences for metabolic efficiency of the whole organism and its competitiveness in an ecosystem, so there must presumably be a selective advantage for secreting high-cost AAs into the public commons. In animals, costly signals are associated with generation of 'trust' in the honesty of a signal (Polnaszek and Stephens 2013), and this phenomenon even helps prevent cheating between cells within the animal (Krakauer and Pagel 1996). By analogy, this raises the possibility that the costliness of the A-signal may improve the control of cheating genotypes and/or competitors of the *M. xanthus* multicellular population. For instance, if cheats are present and reducing the amount/quality of A-signal, development may be aborted or precluded. Having a costly A-signal might be expected to make A-signalling more sensitive to the presence of cheats. It may also be important in preventing the population from being 'tricked' into inappropriate development by secretion of AA by competitors/cheats.

The value of AAs as a food source is possibly important in precluding the widespread adoption of AAs as extracellular signals, leaving the mode of communication secure for myxobacterial communication. This uniqueness is likely to be important for a predatory organism, as otherwise it would not be able to distinguish between signals originating from prey, or from other members of its colony. It is possible that the A-signal represents an extracellular source of nutrients for fuelling the developmental process; however, this is robustly argued against by Kuspa et al. (1992a). Later in development, a subset of cells is known to autolyse to allow continued development by the survivors (Wireman and Dworkin 1977), but the AA released at that point would be expected to be more 'balanced' than those comprising the A-signal. It is likely that *M. xanthus* signals using a subset of AA rather than other costly food molecules, as its primary food source is proteins, obviating any cost in producing the signal, and AAs are small water-soluble compounds with diverse chemical properties.

AACB varies within an organism's sub-proteomes

To relate AACB of subsets of genes within an organism to aspects of its ecology/lifestyle, requires definition of subsets of genes/proteins involved in particular processes and lifestyle/life cycle changes (Elser et al. 2011).

In our model organism *M. xanthus*, distinct subsets of proteins involved in different biological processes and with different sub-cellular localisation have differing AACB. Presumably, this is due to selective advantages, shaped by cost, availability, auxotrophy and other factors (such as functional roles). Such patterns of AACB are not apparent from looking at the whole organism, but could provide insights into the selective advantages/pressures operating under different stages of the life cycle and in different environmental niches.

Conflict of interest The authors declare that they have no conflict of interest.

References

Akashi H, Gojobori T (2002) Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. Proc Natl Acad Sci USA 99:3695–3700

Berleman JE, Allen S, Danielewicz MA, Remis JP, Gorur A, Cunha J, Hadi MZ, Zusman DR, Northen TR, Witkowska HE, Auer M (2014) The lethal cargo of *Myxococcus xanthus* outer membrane vesicles. Front Microbiol 5:474

Bohlin J (2011) Genomic signatures in microbes—properties and applications. Sci World J 11:715–725

Bragg JG, Wagner A (2006) Protein carbon content evolves in response to carbon availability and may influence the fate of duplicated genes. Proc R Soc B 274:1063–1070

Bretscher AP, Kaiser D (1978) Nutrition of *Myxococcus xanthus*, a fruiting myxobacterium. J Bacteriol 133:763–768

Chen W, Shao Y, Chen F (2013) Evolution of complete proteomes: guanine-cytosine pressure, phylogeny and environmental influences blend the proteomic architecture. BMC Evol Biol 13:219

Curtis PD, Shimkets LJ (2008) Metabolic pathways relevant to predation, signaling, and development. In: Whitworth DE (ed) Myxobacteria: multicellularity and differentiation. ASM Press, Washington, DC, pp 241–258

Dahl JL, Tengra FK, Dutton D, Yan J, Andacht TM, Coyne L, Windell V, Garza AG (2007) Identification of major sporulation proteins of *Myxococcus xanthus* using a proteomic approach. J Bacteriol 189:3187–3197

Elser JJ, Acquisti C, Kumar S (2011) Stoichiogenomics: the evolutionary ecology of macromolecular elemental composition. Trends Ecol Evol 26:38–44

Evans AR, Whitworth DE (2010) Correlations between the role, sequence conservation, genomic location and severity of phenotype in myxobacterial developmental genes. FEMS Microbiol Lett 312:40–45

Evans AGL, Davey HM, Cookson A, Currinn H, Cooke-Fox G, Stanczyk P, Whitworth DE (2012) Predatory activity of *Myxococcus xanthus* outer membrane vesicles and properties of their hydrolase cargo. Microbiology 158:2742–2752

Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson BØ (2007) A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol Syst Biol 3:121

Foerstner KU, von Mering C, Hooper SD, Bork P (2005) Environments shape the nucleotide composition of genomes. EMBO Rep 6:12

Galka F, Wai SN, Kusch H, Engelmann S, Hecker M, Schmeck B, Hippenstiel S, Uhlin BE, Steinert M (2008) Proteomic characterization of the whole secretome of *Legionella pneumophila* and functional analysis of outer membrane vesicles. Infect Immun 76:1825–1836

Hershberg R, Petrov DA (2010) Evidence that mutation is universally biased towards AT in bacteria. PLoS Genet 6:e1001115

Hildebrand F, Meyer A, Eyre-Walker A (2010) Evidence of selection upon genomic GC-content in bacteria. PLoS Genet 6:e1001107

Huang Y, Niu B, Gao Y, Fu L, Li W (2010) CD-HIT Suite:a web server for clustering and comparing biological sequences. Bioinformatics 26:680–682

Kahnt J, Aguiluz K, Koch J, Treuner-Lange A, Konovalova A, Huntley S, Hoppert M, Søgaard-Andersen L, Hedderich R (2010) Profiling the outer membrane proteome during growth and development of the social bacterium *Myxococcus xanthus* by selective biotinylation and analyses of outer membrane vesicles. J Proteome Res 9:5197–5208

Krakauer DC, Pagel M (1996) Selection by somatic signals: the advertisement of phenotypic state through costly intercellular signals. Philos Trans R Soc Lond B Biol Sci 351:647–658

Kuspa A, Plamann L, Kaiser D (1992a) Identification of heat-stable A-factor from *Myxococcus xanthus*. J Bacteriol 174:3319–3326

Kuspa A, Plamann L, Kaiser D (1992b) A-signalling and the cell density requirement for *Myxococcus xanthus* development. J Bacteriol 174:7360–7369

Mann S, Chen YP (2010) Bacterial genomic G + C composition-eliciting environmental adaptation. Genomics 95:7–15

Markowitz VM, Chen IM, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Jacob B, Huang J, Williams P, Huntemann M, Anderson I, Mavromatis K, Ivanova NN, Kyrpides NC (2012) IMG: the Integrated Microbial Genomes database and comparative analysis system. Nucl Acids Res 40 (Database issue):D115–22

McCleary WR, Esmon B, Zusman DR (1991) *Myxococcus xanthus* protein C is a major spore surface protein. J Bacteriol 173:2141–2145

Merchant SS, Helmann JD (2012) Elemental economy: microbial strategies for optimizing growth in the face of nutrient limitation. Adv Microb Physiol 60:91–210

Patel NA, Crombie A, Slade SE, Thalassinos K, Hughes C, Connolly JB, Langridge J, Murrell JC, Scrivens JH (2012) Comparison of one- and two-dimensional liquid chromatography approaches in the label-free quantitative analysis of *Methylocella silvestris*. J Proteome Res 11:4755–4763

Polnaszek TJ, Stephens DW (2013) Why not lie? Costs enforce honesty in an experimental signalling game. Proc Biol Sci 281:20132457

Rahman O, Cummings SP, Harrington DJ, Sutcliffe IC (2008) Methods for the bioinformatic identification of bacterial lipoproteins encoded in the genomes of Gram-positive bacteria. World J Microbiol Biotech 24:2377–2382

Rosenberg E, Keller KH, Dworkin M (1977) Cell density-dependent growth of *Myxococcus xanthus* on casein. J Bacteriol 129:770–777

Velicer GJ, Vos M (2009) Sociobiology of the myxobacteria. Annu Rev Microbiol 63:599–623

Whitworth DE (2008) Myxobacteria: multicellularity and differentiation. ASM Press, Washington, DC

Whitworth DE (2011) Myxobacterial vesicles: death at a distance? Adv Appl Microbiol 75:1–31

Wireman JW, Dworkin M (1977) Developmentally induced autolysis during fruiting body formation by *Myxococcus xanthus*. J Bacteriol 129:798–802

Zundel MA, Basturea GN, Deutscher MP (2009) Initiation of ribosome degradation during starvation in *Escherichia coli*. RNA 15:977–983