No Evidence That Nitrogen Limitation Influences the Elemental Composition of Isopod Transcriptomes and Proteomes

Clémentine M. Francois,[1] Laurent Duret,[2] Laurent Simon,[1] Florian Mermillod-Blondin,[1] Florian Malard,[1] Lara Konecny-Dupré,[1] Rémi Planel,[2] Simon Penel,[2] Christophe J. Douady,[1,3] and Tristan Lefébure*,[1]

[1]Univ Lyon, Université Claude Bernard Lyon 1, CNRS, ENTPE, Laboratoire d'Ecologie des Hydrosystèmes Naturels et Anthropisés UMR5023, Villeurbanne, France

[2]Univ Lyon, Université Claude Bernard Lyon 1, CNRS, Laboratoire de Biométrie et Biologie Evolutive UMR5558, Villeurbanne, France

[3]Institut Universitaire de France, Paris, France

***Corresponding author:** E-mail: tristan.lefebure@univ-lyon1.fr.

Associate editor: Deepa Agashe

Abstract

The field of stoichiogenomics aims at understanding the influence of nutrient limitations on the elemental composition of the genome, transcriptome, and proteome. The 20 amino acids and the 4 nt differ in the number of nutrients they contain, such as nitrogen (N). Thus, N limitation shall theoretically select for changes in the composition of proteins or RNAs through preferential use of N-poor amino acids or nucleotides, which will decrease the N-budget of an organism. While these N-saving mechanisms have been evidenced in microorganisms, they remain controversial in multicellular eukaryotes. In this study, we used 13 surface and subterranean isopod species pairs that face strongly contrasted N limitations, either in terms of quantity or quality. We combined in situ nutrient quantification and transcriptome sequencing to test if N limitation selected for N-savings through changes in the expression and composition of the transcriptome and proteome. No evidence of N-savings was found in the total N-budget of transcriptomes or proteomes or in the average protein N-cost. Nevertheless, subterranean species evolving in N-depleted habitats displayed lower N-usage at their third codon positions. To test if this convergent compositional change was driven by natural selection, we developed a method to detect the strand-asymmetric signature that stoichiogenomic selection should leave in the substitution pattern. No such signature was evidenced, indicating that the observed stoichiogenomic-like patterns were attributable to nonadaptive processes. The absence of stoichiogenomic signal despite strong N limitation within a powerful phylogenetic framework casts doubt on the existence of stoichiogenomic mechanisms in metazoans.

Key words: **stoichiogenomics, RNA and protein composition, nitrogen cost, orthologous gene families, comparative approach, N availability, C:N mismatch.**

Downloaded from http://mbe.oxfordjournals.org/ at INIST-CNRS on September 29, 2016

Article

Introduction

A substantial amount of the cell energy budget is invested in proteins and RNAs synthesis (up to 75%; Carter and Houlihan 2001; Wagner 2005; Lane and Martin 2010; Flamholz et al. 2014), a function sustained by a set of biochemical pathways among the most conserved in living beings (Karp et al. 2005). When essential nutrients become limiting in the environment of cells or organisms, adaptive responses have been observed at multiple levels ranging from genome streamlining (Hessen et al. 2010) to the development of alternative metabolic pathways (Merchant and Helmann 2012). At lower levels of organization, nutrient limitations might have an influence on the elemental composition of genomes, transcriptomes, and proteomes. Indeed, the 20 amino acids, as well as the 4 nt, differ in the number of key elements they contain [e.g., carbon (C), nitrogen (N), oxygen (O), or sulfur (S)]. Thus, changes in the elemental composition of DNA, RNAs, or proteins may to some extent decrease nutrient requirements and increase the fitness of an organism facing environmental limitations (this emerging field of study is referred to as "stoichiogenomics" in the literature; Elser et al. 2011). Studies in yeast showed that even a single amino acid change can be visible to natural selection on the basis of its C, N, or S cost (Bragg and Wagner 2009). The influence on the fitness is expected to be stronger when changes happen in highly expressed (HE) proteins (Bragg and Wagner 2009) but also in proteins involved in nutrient assimilation pathways, allowing these proteins to be precisely synthesized under nutrient limitation (Baudouin-Cornu et al. 2001).

Nitrogen is an essential and frequently limiting factor of ecosystem's primary productivity (Vitousek and Howarth 1991; Karl et al. 2002; Elser et al. 2007). While N is a major constituent of cells (~7–10% of dry weight) and makes up 17% and 14.5% of proteins and nucleic acids, respectively (Sterner and Elser 2002), only a few specialized N-fixing organisms can directly assimilate N from the large pool of atmospheric N_2. Many organisms may therefore be constrained by N limitation to meet their metabolic requirements. The 4 nt represent different N-costs: thymine (T) as well as uracile (U) requires two N atoms, cytosine (C) three, and adenine (A)

© The Author 2016. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

and guanine (G) five. Likewise, all 20 amino acids contain one constitutive N atom in their amine functional group but only six of them require additional N atoms in their side chain (one additional N atom for asparagine, glutamine, and lysine; two for histidine and tryptophan; three for arginine). The genetic code is redundant and amino acid substitutions can be functionally equivalent depending on their physicochemical properties. Within the theoretical framework of stoichiogenomics, it is therefore expected that a long-standing N limitation would select for the preferential use of functionally equivalent but N-poor amino acids and nucleotides. In metazoans, the total pool of nitrogen is typically allocated at 80% to the proteins, 17% to the RNAs, and 3% to the DNA and various metabolites (Sterner and Elser 2002; Larsen et al. 2011). Stoichiogenomic compositional changes associated to N limitation should therefore be mostly concentrated in the elemental composition of the proteome and transcriptome, and within these two largest N pools, stoichiogenomic selection should favor compositional changes in the most HE proteins and mRNAs (Bragg and Wagner 2009).

The stoichiogenomic theory has been based on the assumption that nutrient limitations can influence the elemental composition of proteins and nucleic acids (Elser et al. 2011). An environment is considered limiting if the nutrient composition of the available food sources limits the consumer performance. N limitation is demonstrable when a substantial N addition increases the organism's growth or primary production (Vitousek and Howarth 1991; Downing et al. 1999; Elser et al. 2007; Moore et al. 2013). However, such supplementation experiments are highly challenging to complete in situ. As a result, to our knowledge, N limitation was never stricto sensu demonstrated in the stoichiogenomic studies published so far. Instead, most of these studies actually relied on measures of environmental N quantity as a proxy for N limitation (e.g., nitrate concentrations in Grzymski and Dussaq 2012). While this proxy is admittedly imperfect, some articles indeed support a link between nutrient quantity and limitation (e.g., in Moore et al. 2013). However, approximating N limitation solely by N quantity is testing only one side of the story as resource quality can also induce limitation. While resource quality has been largely disregarded in the stoichiogenomic studies, it can be estimated by the mismatch between the elemental composition of the consumer and that of its resources (Sterner and Elser 2002), a strong mismatch indicating a poor quality of the resources for the consumer (i.e., strong N limitation in quality).

Historically, the first reports of compositional changes linked to nutrient limitations in terms of quantity date back three decades ago (Cuhel et al. 1981; Mazel and Marliere 1989). Yet, the stoichiogenomic literature has been thriving recently with many studies providing robust evidence of N-saving mechanisms in micro-organisms facing harsh conditions. These N-saving mechanisms have been identified at genomic (Luo et al. 2015) and proteomic (Lv et al. 2008; Grzymski and Dussaq 2012) levels in marine micro-organisms inhabiting N-depleted environments. As expected, N-saving mechanisms were stronger in HE proteins (Li et al. 2009; Gilbert and Fagan 2011; Grymski and Dussaq 2012) and in nitrogen stress response proteins (upregulated in situation of N limitation; Gilbert and Fagan 2011).

While micro-organisms provide evidence of N-savings under N limitation in quantity, this issue remains an open question in multicellular eukaryotes for which the literature is scarce. For example, the N-content of transcriptome and proteome of wild plants have been reported to be lower than that of domesticated plants (Acquisti et al. 2009). This difference has been interpreted as an evidence of relaxed selection for N-savings in domesticated plants supplemented by N-rich fertilizers as opposed to wild plants (Acquisti et al. 2009). The strength of these N-saving mechanisms was also reported to increase with protein expression level in plants and insects (Elser et al. 2006; Gilbert et al. 2013). However, studies reporting N-saving mechanisms at transcriptomic to proteomic levels among multicellular eukaryotes were criticized for several reasons (e.g., Gunther et al. 2013). They did not consider the phylogenetic relationships between the studied species (e.g., see Elser et al. 2006), they did not measure the actual N quantity available in the environment (e.g., see Gilbert et al. 2013), and none of them considered resource quality. Comparisons of sequence elemental composition were also made between sequence bulks and not between orthologous sequences. Finally, the influence of nonadaptive processes, such as changes of mutational pattern or variation of the strength of GC-Biased Gene Conversion (gBGC) which may both create stoichiogenomic-like patterns, were not considered (Gunther et al. 2013). In addition to these methodological issues, most studies focused on plants (Elser et al. 2006; Acquisti et al. 2009) whose nutrition relies mainly on inorganic N uptake from their environment while metazoans obtain N through the amino acids contained in their diet. This difference implies that the sparse evidence in the literature for N-savings in plants may not apply to metazoans. All these shortcomings call for a strict test of the stoichiogenomic theory in metazoans, based on a comparative framework covering different species facing contrasted N limitation.

To challenge the stoichiogenomic theory in metazoans, we took advantage of the very low nutrient quantity that prevails in most groundwater (GW) habitats compared with surface water (SW) habitats (Poulson and Lavoie 2000; Gibert and Deharveng 2002; Venarsky et al. 2014). The trophic resources available in GW are expected to be of low quality as organic matter entering GW habitats has been depleted from its most biodegradable fraction during its transfer from surface to GW (Poulson and Lavoie 2000; Simon et al. 2003; Mermillod-Blondin et al. 2015). GW habitats are thus expected to represent strong N limitation in terms of quantity and quality. GW has been repeatedly colonized by isopod species of the Asellidae family (Morvan et al. 2013), thereby providing independent replicates of the ecological transition to likely N-limited habitats. These multiple ecological transitions between sister species of metazoans over a relatively short time-scale are among the harshest shifts recorded on Earth. We built a comparative framework within the Asellidae family by defining 13 phylogenetically independent species pairs (sensu Felsenstein 1985), each pair being composed of one SW and one GW species sharing a common SW ancestor (fig.

Downloaded from http://mbe.oxfordjournals.org/ at INIST-CNRS on September 29, 2016

1). Within this framework, we estimated for each species the N limitation in terms of quantity by measuring the quantity of N contained in the available trophic resources edible by asellids (hereafter referred to as "N availability"). We also estimated the N limitation in terms of quality by determining the mismatch between the carbon-to-nitrogen molar ratio (C:N) of the isopods and that of their resources (hereafter "C:N mismatch"). N-saving compositional changes were tracked by sequencing the trancriptome of these 26 species then inferring their proteome via in silico translation. Under the stoichiogenomic theory, we expected N-saving compositional changes, primarily in the proteomes, but also in the transcriptomes of GW species that are facing strong N-limitation, either in terms of quantity or quality. Through a set of complementary analyses, we tested for a change in N-use at two levels: 1) at the global proteome and transcriptome level, and 2) at the level of orthologous gene families. While the former best captures changes in expression levels that could impact the overall N budget, the latter is better at detecting changes in the substitution pattern that are consistent with the stoichiogenomic theory. Several nonadaptive forces such as the gBGC or modifications of the mutational pattern can also generate analogous compositional changes. To further test for N-saving compositional changes, we took advantage of the fact that these changes are strand-specific and developed a new method to test for N-saving mechanisms in transcriptomes. No evidence of N-saving was found in the total N-budget of transcriptome or proteome, in the expression of the most N-costly transcripts or proteins, in the average N-cost of a transcript or protein, in the N-cost of substitutions, nor in the pattern of strand asymmetry.

Results

Ecological Parameters

This study combined both genomic and ecological data. The two proxies for N limitation (in terms of quantity and quality) were estimated in the habitat of 18 isopod species (nine SW and nine GW sites). N availability, corresponding to the quantity of N contained in all available trophic resources known to be edible by asellids, was determined in each of these 18 sampling sites. The average N availability was estimated at 2.62 and 0.83 g of N . m^{-2} for SW and GW habitats, respectively (table 1, supplementary table S1, Supplementary Material online). As expected, GW habitats were significantly depleted in N compared with SW habitats (Wilcoxon test, P value = 0.015). The C:N mismatch between the available resources known to be edible by asellids and the focal isopods was used as a proxy for N limitation in terms of resource quality. The focal habitats spanned a wide range of C:N mismatches (from −1.8 to 41.4, table 1). The C:N mismatch showed no significant difference between SW and GW habitats (Wilcoxon test, P value = 0.094), though we observed a trend for relatively greater C:N mismatch in SW. Besides, the C:N mismatch was not correlated to N availability (t-test, P value = 0.724). We subsequently performed two sets of analyses: "global" analyses in which all isopod species are considered, and finer "within-pair" analyses in which we compare the orthologous genes between both species of each pair. In global analyses, the influence of the habitat transition will be tested using three ecological parameters: a SW versus GW categorical parameter (habitat type), the N availability (proxy for N limitation in terms of quantity), and the C:N mismatch (proxy for N limitation in terms of quality). For within-pair analyses, since the habitat type was partially correlated to the N availability and C:N mismatch, the "polarization" of the stoichiogenomic hypothesis (the most N-limited species in the pair) was defined in three different ways: 1) according to the habitat type (GW more N-limited than SW; "habitat" hypothesis); 2) according to the N limitation in quantity (N availability; "quantity" hypothesis); and 3) according to the N limitation in quality (C:N mismatch; "quality" hypothesis). These polarization rules are explicited for each pair in the supplementary table S2, Supplementary Material online.

Bulk Transcriptome and Proteome N Budget

We estimated the transcriptome N budget by weighting each coding sequence (CDS; supplementary table S3, Supplementary Material online) by its expression level, allowing longer and HE genes to have a stronger impact on the total N usage. This total N budget is normalized for 1 million transcripts and corresponds to the Relative number of N Atoms per Transcriptome (RNAT, supplementary table S4, Supplementary Material online). The same reasoning can be applied at the proteome level (Relative number of N Atoms per Proteome—RNAP, supplementary table S4, Supplementary Material online). In absence of direct proteome quantification, RNAP was computed using mRNA expression levels as a proxy for protein abundances (see M&M). Recent studies showed that mRNA abundances are strongly correlated to protein abundances (Csardi et al. 2015; Li and Biggin 2015). Thus, though indirect, the RNAP measure should allow the comparison of overall protein N budget. Neither the RNAP nor the RNAT were significantly different between SW and GW (table 2). Likewise, neither RNAP nor RNAT were significantly correlated with the N availability or C:N mismatch (table 2).

We also tested if CDS and proteins of GW species tended to contain less N atoms, independently of their expression. The average N-cost per site was estimated, for proteins, as the number of N Atoms per Residue Side Chain (NARSC, as defined in Acquisti et al. 2009 and Baudouin-Cornu 2001). The NARSC was not correlated with the habitat type, N availability or C:N mismatch (fig. 2a, table 2, supplementary table S4, Supplementary Material online). The same results were found even when the analysis was restricted to the 10% most-expressed proteins where selection pressure for N-savings is expected to be stronger (PGLS [phylogenetic generalized least-squares method, Martins and Hansen 1997], P value = 0.267, 0.313, and 0.668 for the influence of habitat, N availability and C:N mismatch, respectively). All species displayed a significant NARSC decrease in its 10% most-expressed proteins (supplementary fig. S1, Supplementary Material online). However, the magnitude of this N-cost reduction was not stronger in GW compared with SW species and was not correlated with N availability or C:N mismatch (PGLS on the ratio [NARSC of the

Downloaded from http://mbe.oxfordjournals.org/ at INIST-CNRS on September 29, 2016

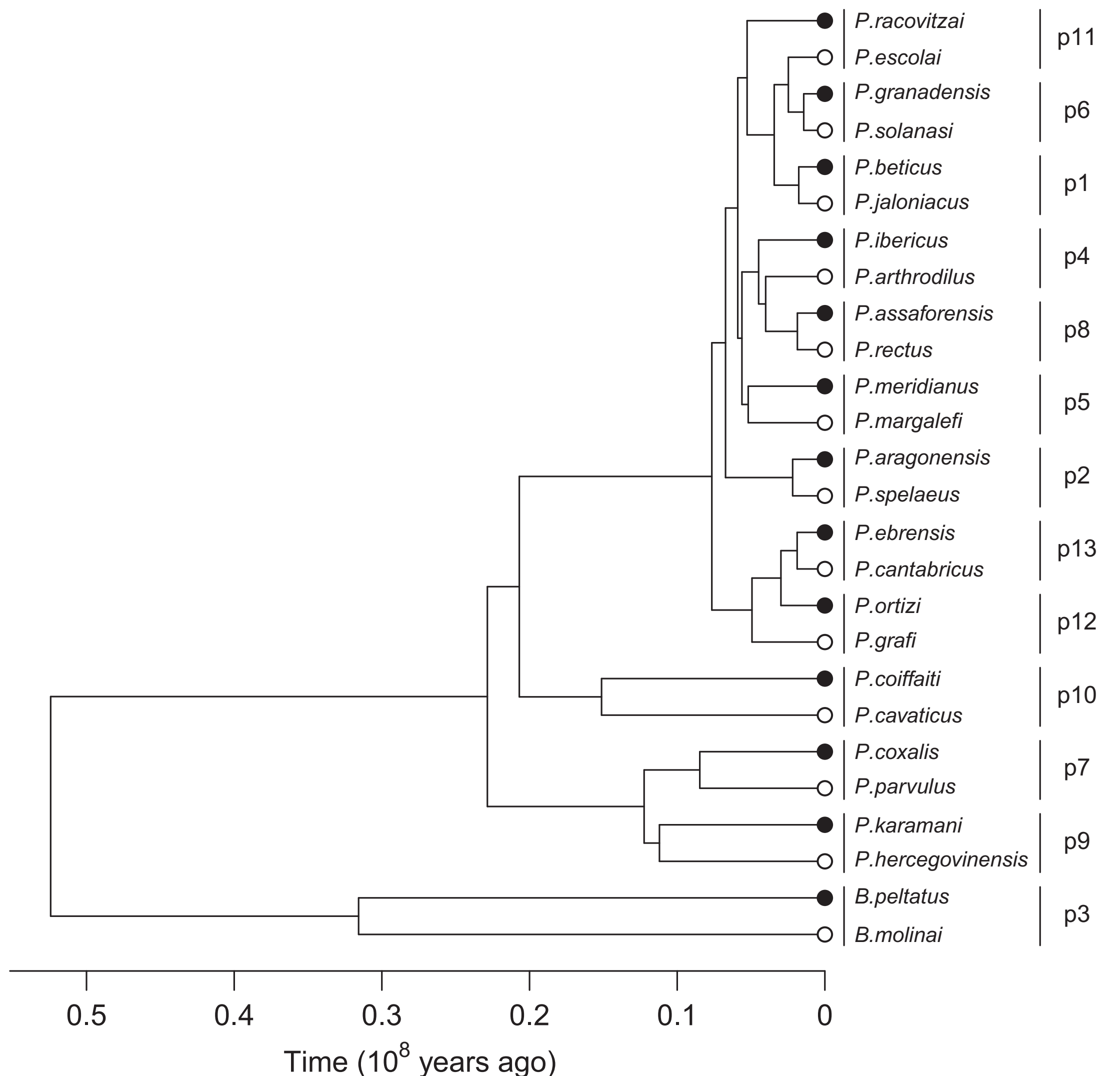

FIG. 1. Chronogram of the 13 pairs of isopod species. Vertical bars next to the tree indicate species pairs composed of one SW (surface water, black circles) and one GW (groundwater, white circles) species. Pair numbers as in table 1.

10% least-expressed/NARSC of the 10% most-expressed proteins], P value = 0.194, 0.202, and 0.240 for habitat, N availability or C:N mismatch, respectively).

We also analyzed specifically the third codon positions of CDS, as most substitutions at these sites are synonymous and are thus more likely to display a stoichiogenomic signal, if any. The average N-cost per site was estimated as the number of N Atoms per Base in third position of the codon (NAB3), ranging from 2 (T/U) to 5 (A and G). The NAB3 was significantly higher in SW than in GW species but was not significantly correlated with the N availability or C:N mismatch (table 2 and fig. 2b, supplementary table S4, Supplementary Material online).

Variation of the Substitution N-Cost within Families of Orthologous Genes

One possible limitation of the above bulk analyses is that divergence between SW and GW species might be too recent for selection to leave a detectable transcriptome or proteome-wide N-usage change. To circumvent this, we identified orthologous genes (supplementary table S3, Supplementary Material online) and analyzed the pattern of substitutions that have occurred since the divergence between SW and GW species, and hence that should have been subjected to different selective pressures on N-usage in the two lineages. We focused on substitutions which changed the N-cost per site. GW species display a set of life-history traits that are likely to slow down their substitution rates (e.g., lower metabolic rate and longer generation time). Therefore, a net comparison of the number of N-costly substitutions across taxa might be systematically biased as SW will tend to display more substitutions and thus more N-costly changes. To this end, we calculated the proportion of substitutions that increased the N-cost (supplementary table S3, Supplementary Material online), which has the advantage to be insensitive to differences in substitution rates between GW and SW species. We estimated this proportion in proteins and third codon positions (named p_{AA} and p_{B3}, respectively; supplementary table S4, Supplementary Material online).

In proteins, a global test including the 13 species pairs showed that p_{AA} was not significantly correlated with the

Downloaded from http://mbe.oxfordjournals.org/ at INIST-CNRS on September 29, 2016

Table 1. Ecological Parameters for the 26 Isopods Species.

Species	Code	Habitat	Pair	N Availability ($g.m^{-2}$)	C:N Mismatch
Proasellus beticus	Pbeti	SW	p1	0.87	41.42
Proasellus jaloniacus	Pjalo	GW		0.21	11.92
Proasellus aragonensis	Parag	SW	p2	NA	NA
Proasellus spelaeus	Pspel	GW		0.39	6.47
Bragasellus peltatus	Bpelt	SW	p3	4.15	10.53
Bragasellus molinai	Bmoli	GW		0.19	7.58
Proasellus ibericus	Piber	SW	p4	2.05	19.55
Proasellus arthrodilus	Parth	GW		0.18	−0.52
Proasellus meridianus	Pmeri	SW	p5	3.24	9.08
Proasellus margalefi	Pmarg	GW		NA	NA
Proasellus granadensis	Pgran	SW	p6	0.61	6.69
Proasellus solanasi	Psola	GW		3.48	3.95
Proasellus coxalis	Pcoxa	SW	p7	2.39	8.59
Proasellus parvulus	Pparv	GW		0.53	9.65
Proasellus assaforensis	Passa	SW	p8	0.24	4.12
Proasellus rectus	Prect	GW		1.58	−1.79
Proasellus karamani	Pkara	SW	p9	9.17	7.00
Proasellus hercegovinensis	Pherc	GW		0.24	11.04
Proasellus coiffaiti	Pcoif	SW	p10	0.86	19.23
Proasellus cavaticus	Pcava	GW		0.65	0.50
Proasellus racovitzai	Praco	SW	p11	NA	NA
Proasellus escolai	Pesco	GW		NA	NA
Proasellus ortizi	Porti	SW	p12	NA	NA
Proasellus grafi	Pgraf	GW		NA	NA
Proasellus ebrensis	Pebre	SW	p13	NA	NA
Proasellus cantabricus	Pcant	GW		NA	NA

NOTE.—The C:N mismatch corresponds to the mismatch between the C:N of the available food sources and the C:N of isopods (see text for details). "NA" indicates that the quantification of trophic resources was not achievable.

habitat type, N availability or C:N mismatch (fig 2c and table 2). The trend for a positive correlation between p_{AA} and C:N mismatch (table 2) was opposite to stoichiogenomic prediction. We also tested the N-saving stoichiogenomic theory within each species pair through a bootstrap procedure. Whether N limitation is estimated in terms of habitat, N availability (quantity) or C:N mismatch (quality), none or only one pair displayed the expected stoichiogenomic-like pattern (p4; table 3 and supplementary table S5, Supplementary Material online). We then restricted the analysis to the HE (20% most-expressed) gene families where selection for N-savings is expected to be stronger. Results were roughly the same in the HE genes, whether in global (PGLS, *P* value = 0.523, 0.548, and 0.167 for habitat, N availability and C:N mismatch, respectively) or within-pair tests (table 3 and supplementary table S5, Supplementary Material online).

At third codon positions, a global test including the 13 species pairs showed that SW species had a significantly higher p_{B3} than GW species (fig 2d and table 2). p_{B3} was positively correlated with N availability but showed no correlation with C:N mismatch (table 2). We also tested the N-saving stoichiogenomic theory within each species pair through a bootstrap procedure. When N limitation was estimated by the habitat type or by the N availability (quantity hypothesis), we observed some stoichiogenomic-like signal (6 pairs out of 13 and 4 pairs out of 8, respectively; table 3 and supplementary table S5, Supplementary Material online). When N limitation was estimated by the C:N mismatch (quality hypothesis), only one pair displayed the expected stoichiogenomic-like pattern (p6; supplementary table S5, Supplementary Material online). These patterns might result from stoichiogenomic selection in response to N limitation in quantity in GW habitats. However, five of the six pairs displaying a stoichiogenomic-like pattern (p_{B3} in SW > p_{B3} in GW) also displayed a lower %GC at third codon positions (GC3) in the GW species (table 3). This stoichiogenomic-like pattern (of lower p_{B3} in GW species) may thus result from nonadaptive processes such as a decrease in gBGC strength in GW species.

Testing for Strand Asymmetries in Substitution Rates

A pertinent way to discriminate the effects of adaptive (stoichiogenomic) and nonadaptive processes is to look at the strand asymmetries in substitution rates (see the corresponding paragraph in the M&M). N-savings are expected to leave a transcriptome wide asymmetric substitution footprint. Indeed, between two complementary substitutions, stoichiogenomic selection should favor the substitution reducing the N-cost per base on the coding strand (e.g., A→T over T→A and G→C over C→G), generating strand asymmetries in substitution rates (e.g., qA→T > qT→A). We estimated the level of asymmetry at third codon positions for these two transversion classes for the 26 species using $\Delta AT = (qA{\rightarrow}T)-(qT{\rightarrow}A)$ and $\Delta GC = (qG{\rightarrow}C)-(qC{\rightarrow}G)$. These two measures are expected to increase in N-limited species (e.g., $\Delta_{GW} > \Delta_{SW}$ for the habitat hypothesis; supplementary table S6, Supplementary Material online).

All species displayed some level of substitutional strand asymmetry with $\Delta AT > 0$ and $\Delta GC < 0$ (fig. 3a and *b*). A global test including the 13 pairs showed no significant difference of ΔAT and ΔGC between SW and GW species (PGLS, *P* value = 0.592 and 0.594 for ΔAT and ΔGC, respectively). Similarly, ΔAT and ΔGC were not correlated to N availability (PGLS, *P* value = 0.242 and 0.161 for ΔAT and ΔGC, respectively) or C:N mismatch (PGLS, *P* value = 0.907 and 0.900 for ΔAT and ΔGC, respectively). Within each species pair, the stoichiogenomic theory was tested through a bootstrap procedure. Whether N limitation is estimated in terms of habitat, N availability (quantity) or C:N mismatch (quality), only two or three pairs displayed the expected stoichiogenomic-like pattern for ΔAT and one or two pairs for ΔGC (table 4 and supplementary table S7, Supplementary Material online). Only one of these tests remained significant when corrected for multiple hypotheses testing (supplementary table S7, Supplementary Material online). No pair displayed the expected stoichiogenomic-like patterns both on ΔAT and ΔGC.

Expression of the Most N-Costly Gene Families

One stoichiogenomic expectation that does not require substitutions to accumulate and might therefore be faster to achieve is a reduction of the total N-budget through a decrease of the expression of the most N-costly transcripts and proteins. While we did not observe any variation of the global transcriptome and proteome N-budget (RNAT and RNAP) linked to N limitation, this global approach might not be

Downloaded from http://mbe.oxfordjournals.org/ at INIST-CNRS on September 29, 2016

Table 2. Global Analyses (phylogenetic generalized least-squares models) Testing the Effect of Ecological Parameters on the N-Cost of the Proteome/Transcriptome.

	PROTEOME					TRANSCRIPTOME				
	Dependent Variable	Explanation Variable	LRT *P* Value	Coefficient	R^2	Dependent Variable	Explanation Variable	LRT *P* Value	Coefficient	R^2
Bulk analysis	RNAP	Habitat	0.584	20,379,020	0.011	RNAT	Habitat	0.600	165,817,835	0.011
		Quantity (N availability)	0.853	−3,206,338	0.002		Quantity (N availability)	0.871	−23,730,695	0.001
		Quality (C:N mismatch)	0.783	−6,629,593	0.004		Quality (C:N mismatch)	0.773	−58,882,984	0.005
	NARSC	Habitat	0.379	0.00071	0.029	NAB3	<u>Habitat</u>	<u>0.016</u>	<u>−0.00433</u>	<u>0.202</u>
		Quantity (N availability)	0.923	−0.00005	0.001		Quantity (N availability)	0.385	0.00168	0.041
		Quality (C:N mismatch)	0.492	−0.00045	0.026		Quality (C:N mismatch)	0.969	0.00007	< 0.001
Orthologous genes	p_{AA}	Habitat	0.851	0.00492	0.001	p_{B3}	<u>Habitat</u>	<u>0.040</u>	<u>−0.04916</u>	<u>0.150</u>
		Quantity (N availability)	0.966	−0.00051	< 0.001		<u>Quantity (N availability)</u>	<u>0.015</u>	<u>0.03371</u>	<u>0.279</u>
		Quality (C:N mismatch)	0.074	0.02904	0.162		Quality (C:N mismatch)	0.160	0.02800	0.104

Note.—Significant relationships are indicated in underlined. LRT *P* value: likelihood ratio test between the models with and without the given explanation variable. R^2: Magee generalized R^2. Habitat corresponds to the comparison of SW and GW habitats (coefficients are in contrast to the SW habitat). C:N mismatch corresponds to the C:N mismatch between isopods and their food sources. RNAT and RNAP correspond respectively to the Relative number of N Atoms per Transcriptome or its translation into Proteins (normalized for one million transcripts). NARSC and NAB3 correspond respectively to the number of N Atoms per amino acid Residue Side Chain, or per nucleic Base in 3rd position of the codon. p_{AA} and p_{B3} correspond, respectively, to the proportion of N-costly substitutions in proteins and in third codon positions since the last common ancestor of each species pair (i.e., number of inferred substitutions increasing the N-cost/number of substitutions changing the N-cost). See text for details.

sensitive enough to capture subtle gene expression changes. To address this, we performed within-pair analyses to test if the 10% most N-costly orthologous proteins or transcripts displayed lower expression in the most N-limited species of the pair. Whether N limitation is estimated in terms of habitat, N availability or C:N mismatch, only one or two pairs displayed the expected stoichiogenomic pattern, either at the protein or transcript level, with about the same number of pairs displaying the opposite pattern (supplementary table S8, Supplementary Material online). Therefore, we did not capture any consistent expression changes within the most N-costly transcripts and proteins, which is an additional observation that is not consistent with the stoichiogenomic theory.

Discussion

Proxies for N Limitation

Without in situ supplementation experiments, which can hardly be conducted within a comparative metazoan analysis, N limitation is highly challenging to demonstrate stricto sensu and is commonly estimated using ecological proxies. In this study, N limitation was on the one hand approximated in terms of quantity by the N availability which was correlated with the habitat type (lower N availability in GW). On the other hand, N limitation was also approximated in terms of resource quality by the C:N mismatch which was not significantly correlated with the habitat type but showed a trend for greater C:N mismatch in SW. Resources in GW habitats have traditionally been seen as rare and dominated by refractory organic matter of poor quality. Our observations demonstrated that GW resources, while rare, are not necessarily of poor quality. Indeed, food resources in GW habitats tend to be dominated by sedimentary biofilm, which is a high quality resource for isopods ($C:N_{isopods}$ and $C:N_{biofilm}$ ~5–10), whereas SW habitats are often dominated by particulate organic matter (POM) of lower quality ($C:N_{POM}$ >15). In this study, the two proxies for N limitation (in quantity/quality) were not correlated, indicating a decoupling of these two aspects of N limitation in our design. These results support an evolutionary scenario in which GW species have been facing N limitation in terms of quantity since they colonized N-depleted subterranean habitats. GW habitats proved surprisingly variable in terms of resource quality, meaning that we cannot solely rely on the SW/GW contrast when studying N limitation in quality. Nevertheless, based on the literature on aquatic habitats, both proxies for N limitation (in quantity/quality) reached extreme values in our data set, validating the relevance of this Asellidae model to track N-saving stoichiogenomic changes.

No Evidence for N-Saving in Asellidae Transcriptomes and Proteomes in Response to N Limitation

In this study, ecological N limitation has been considered either in terms of quantity (N availability) and quality (C:N mismatch). Contrary to the stoichiogenomic theory, isopods living in N-limited habitats do not display any selective N-saving compositional change in their proteomes. First, the

Downloaded from http://mbe.oxfordjournals.org/ at INIST-CNRS on September 29, 2016

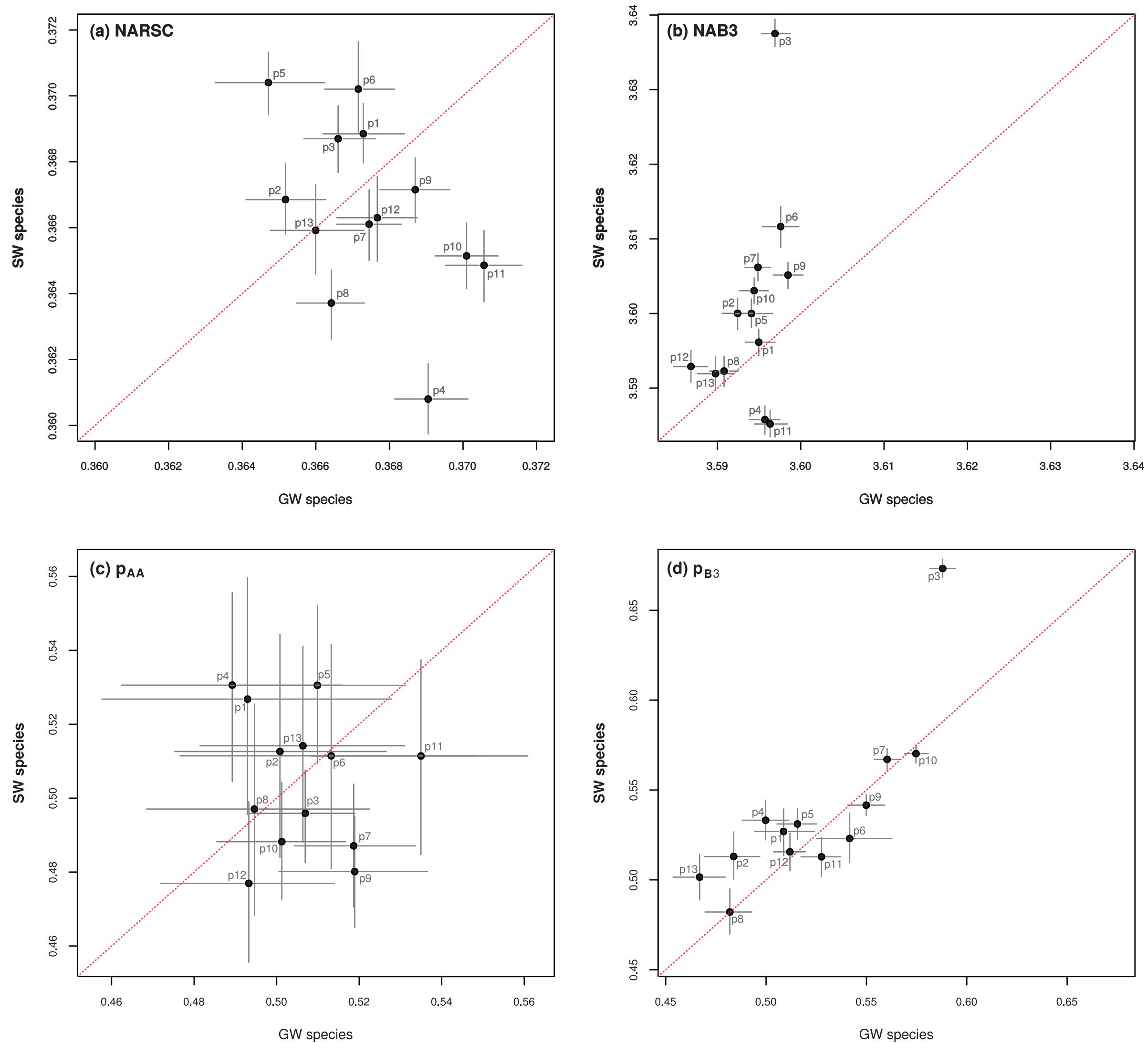

FIG. 2. Comparison of different N-cost estimates between SW and GW species. Each point corresponds to one SW/GW species pair (pair numbers as in table 1) and the bootstrapped 95% confidence intervals are represented for each species. (*a*) NARSC (average number of N Atoms per Residue Side Chain). (*b*) NAB3 (average number of N Atoms per Base in third position of the codon). (*c*) p_{AA} (proportion of N-costly substitutions in proteins). (*d*) p_{B3} (proportion of N-costly substitutions in third codon positions). The *x* axis and *y* axis represent the values for the SW and GW species, respectively. Pairs are thus expected to fall above the bisector line (dotted line).

total N-cost of the proteome normalized per million transcripts was not correlated with the strength of N limitation. Second, the average N-cost per amino acid was not correlated with the strength of N limitation, even when restricting the analysis to the 10% most-expressed proteins which are expected to be under stronger selection pressure. Third, instead of looking at bulk proteomes, we restricted the data set to orthologous genes and estimated for each species the proportion of "N-costly" substitutions since the last common ancestor of each species pair. This proportion of N-costly amino acid substitutions was not correlated with the strength of N limitation, even in the 20% most-expressed protein families. Fourth, the expression of the most N-costly proteins was not reduced in N-limited species.

At the RNA level, we observed that GW species had a significantly lower average N-cost of their third codon positions. Furthermore, the proportion of N-costly substitutions at third codon positions was globally lower in GW species and was correlated with the N limitation in quantity. These observations are consistent with the stoichiogenomic theory. However, these variations could also be caused by nonadaptive processes such as a change in the mutational pattern after the SW/GW divergence or a decrease in the gBGC strength in GW species which might display smaller effective population sizes (Duret and Galtier 2009). This nonadaptive hypothesis is supported by the fact that half of the SW species displayed a higher GC3 than their sister GW species. To disentangle the effect of adaptive and nonadaptive forces acting on third codon positions, we developed a new test of the stoichiogenomic theory, based on the analysis of substitutional strand asymmetries, which is robust to the possible confounding effect of gBGC. Independently of their habitat, all 26 species displayed slight substitutional strand asymmetries: for A↔T transversions, the observed asymmetry was

Downloaded from http://mbe.oxfordjournals.org/ at INIST-CNRS on September 29, 2016

Table 3. Within-Pair Tests of the Stoichiogenomic Hypotheses on p_{AA} and p_{B3} (Habitat).

		Habitat Hypothesis: SW > GW			
		p_{AA}		p_{B3}	GC3
Pair	**Code (SW/GW)**	*All*	*HE*	*all*	*all*
p1	Pbeti/Pjalo	0.081	0.079	0.031	0.352
p2	Parag/Pspel	0.289	0.257	$< 10^{-3}$ *	$< 10^{-3}$ *
p3	Bpelt/Bmoli	0.893	0.558	$< 10^{-3}$ *	$< 10^{-3}$ *
p4	Piber/Parth	0.015	0.008	$< 10^{-3}$ *	$< 10^{-3}$ *
p5	Pmeri/Pmarg	0.087	0.418	0.010 *	0.007 *
p6	Pgran/Psola	0.529	0.224	0.949	0.862
p7	Pcoxa/Pparv	0.995	0.996	0.066	0.150
p8	Passa/Prect	0.441	0.737	0.485	0.077
p9	Pkara/Pherc	1.000	0.978	0.939	$< 10^{-3}$ *
p10	Pcoif/Pcava	0.878	0.924	0.877	0.055
p11	Praco/Pesco	0.901	0.988	0.985	1.000
p12	Porti/Pgraf	0.856	0.631	0.301	0.055
p13	Pebre/Pcant	0.355	0.887	$< 10^{-3}$ *	$< 10^{-3}$ *
		(2)	(3)	(2)	

NOTE.—Here, all within-pair tests were polarised according to the habitat type (GW more N-limited than SW). Are reported the (uncorrected) *P* values of the unilateral bootstrap tests of the stoichiogenomic hypothesis that p_{AA} in SW species > p_{AA} in GW species; and likewise for p_{B3}. Significant relationships are indicated in underlined. "all" correspond to the whole data set of orthologous genes while HE indicates that the data set is restricted to the 20% most-expressed genes. For a given pair, "*" indicates that the test remained significant when corrected for multiple hypothesis testing. Similarly reported are the (uncorrected) *P* values of the unilateral paired Wilcoxon tests of the hypothesis that GC3 in SW species > GC3 in GW species (on the whole data set). The number in brackets at the bottom of the table corresponds to the number of pairs displaying the antistoichiogenomic pattern (SW < GW).

compatible with stoichiogenomic N-savings (ΔAT > 0), but for G$\leftrightarrow$C transversions, the asymmetry (ΔGC < 0) was opposite to the predictions of the stoichiogenomic theory. Moreover, these patterns of substitutional strand asymmetries were not correlated with the strength of N limitation. These observations are therefore not compatible with the stoichiogenomic theory.

In sum, whatever the proxy for N limitation or the N pool considered (proteome or transcriptome), no evidence for N-saving compositional changes was found in the focal isopods.

The Easy Confusion between Adaptive and Nonadaptive Composition Changes

In previous studies describing evidence of stoichiogenomic composition changes in multicellular eukaryotes, the influence of nonadaptive processes have been commonly disregarded. This methodological flaw has already been pointed out by Gunther et al. (2013) who rejected the stoichiogenomic interpretation of base content changes in *Arabidopsis thaliana* (Acquisti et al. 2009), and showed that compositional changes in this species could be explained by mutational bias, purifying selection of functionally deleterious alleles and gBGC alone. In our study, stoichiogenomic-like patterns were observed at few but different levels of the analysis, and particularly at the transcriptome level. Without consideration of alternative nonadaptive processes, these patterns would have been interpreted as evidence for stoichiogenomic mechanisms, while specific tests such as our strand-asymmetry test ruled out such mechanisms. Thus, previously published data sets that were in support of the stoichiogenomic theory warrant a reanalysis that explicitly accounts for nonadaptive processes.

Consistent N-Cost Decrease in HE Proteins

Although unrelated to nutrient limitation, some compositional changes were nevertheless found within the Asellidae proteomes. Indeed, HE proteins consistently displayed lower N-cost per amino acid. This pattern was already described in bacteria (Grzymski and Dussaq 2012), yeast (Li et al. 2009), plants (Elser et al. 2006), and flies (Gilbert et al. 2013) and was often interpreted as an evidence of a physiological selection to preserve the organism's N pool (Elser et al. 2006; Li et al. 2009). However, this pattern was observed in the 26 studied species independently of their habitat. So while this may be an evidence of selection for the optimization of nutrient use, it appears to be disconnected from the nutrient environmental limitation. Alternatively, protein amino acid compositions are known to depend on their function and localization (Pascal et al. 2006). Part of this set of HE proteins may fulfill some particular biological functions requiring specific physicochemical properties for their constitutive amino acids. If this specific group of amino acids tends to be N-poor (e.g., acidic amino acids), this would decrease the N-cost of HE proteins and the observed pattern may just be a functional consequence of their physicochemical properties.

Asellidae as a Relevant Metazoan Model

Absence of stoichiogenomic signal in one metazoan model does not necessarily imply that the stoichiogenomic patterns are absent in metazoans. However, the ecological and phylogenetic characteristics of the Asellidae model suggest that if stoichiogenomic N-saving selection is not acting in this model, it is unlikely to play a significant role in other metazoans. First, mean N availability in our study GW habitats indicated strong N limitation in quantity, in concordance with other GW studies (e.g., Poulson and Lavoie 2000; Venarsky et al. 2014). For example, the standing stock of organic matter (which provides consumers for energy and nutrients) in deep GW habitats was estimated around 10–20 g.m^{-2} (Venarsky et al. 2014), which is one to three order of magnitude lower than values commonly reported for surface streams (68–6,184 g.m^{-2}; data on nine streams, Webster and Meyer 1997). Second, most subterranean species consistently evolved striking biological traits when colonizing GW habitats, such as decreased growth rate and increased starvation resistance (Hervant and Renault 2002; Hüppop 2005). Thus, GW species do live under very low nutrient availabilities and have evolved biological traits in response to it, demonstrating that within this model, environmental limitations which were strong enough to produce important evolutionary changes, did not produce the compositional changes expected under the stoichiogenomic theory. Third, our studied isopods spanned much of the range of C:N mismatches reported for aquatic species (Elser et al. 2000; Cross et al. 2003), including species exhibiting very strong C:N mismatches (up to 41). Thus, some of the studied habitats imposed substantial N limitation in quality for the isopods,

Downloaded from http://mbe.oxfordjournals.org/ at INIST-CNRS on September 29, 2016

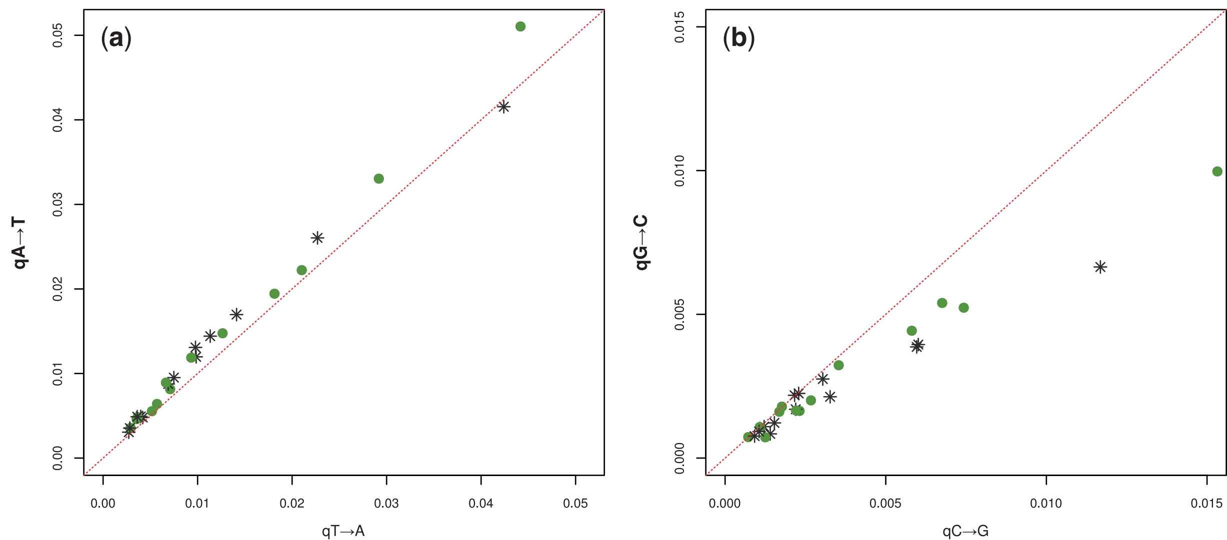

Fig. 3. Parity plots of the complementary substitutions rates. (*a*) Substitution rates from A to T (qA → T) and from T to A (qT → A). (*b*) Substitution rates from G to C (qG → C) and from C to G (qC → G). GW species and SW species, in total of 26 species, are represented by asterisks and circles, respectively. The *y* axis, denoted in bold, corresponds to the substitution which is expected to occur more frequently than its complementary (*x* axis) under the stoichiogenomic prediction of strand asymmetry. Stoichiogenomic selection should move species above the bisector line (dotted line), this deviation being stronger in N-limited species.

Table 4. Within-Pair Tests of the Stoichiogenomic Hypotheses on ΔAT and ΔGC (Habitat).

		Habitat Hypothesis	
Pair	Code (SW/GW)	$\Delta AT_{GW} > \Delta AT_{SW}$	$\Delta GC_{GW} > \Delta GC_{SW}$
p1	Pbeti/Pjalo	0.281	0.836
p2	Parag/Pspel	0.932	0.785
p3	Bpelt/Bmoli	1.000	0.369
p4	Piber/Parth	<u>0.022</u>	0.332
p5	Pmeri/Pmarg	0.480	0.474
p6	Pgran/Psola	0.523	0.928
p7	Pcoxa/Pparv	<u>0.004</u>	0.914
p8	Passa/Prect	0.340	<u>0.029</u>
p9	Pkara/Pherc	0.112	<u>0.004</u>
p10	Pcoif/Pcava	0.736	0.438
p11	Praco/Pesco	0.207	0.834
p12	Porti/Pgraf	<u>0.014</u>	0.072
p13	Pebre/Pcant	0.398	0.761
		(1)	*(0)*

Note.—Here, all within-pair tests were polarized according to the habitat type (GW more N-limited than SW). Are reported the (uncorrected) *P* values of the unilateral bootstrap tests of the stoichiogenomic hypothesis that $\Delta AT_{GW} > \Delta AT_{SW}$ and likewise for ΔGC. Significant relationships are indicated in underlined. No test remained significant when corrected for multiple hypothesis testing. The number in brackets at the bottom of the table corresponds to the number of pairs displaying the antistoichiogenomic pattern (GW < SW).

which in turn produced none of the compositional changes expected under the stoichiogenomic theory. Fourth, even the isopods living in habitats imposing strong N limitation both in quantity and quality (e.g., *P. jaloniacus* and *P. hercegovinensis*) did not display the expected stoichiogenomic changes. Fifth, many stoichiogenomic studies compared taxa much more divergent than the sister species of this study (e.g., Elser et al. 2006), which leads to many more differences between the studied species than in our analysis, possibly reducing our power to detect composition changes. However, the 13 pairs used in this study span a wide range of divergence levels, including some species that accumulated up to 20,000 amino acid inferred changes (p3/p7/p9/p10, see supplementary table S3, Supplementary Material online). On the contrary, the comparison of many sister species within a well-defined phylogenetic framework where species differ in their N limitation but remain close enough so that they still share most of their biological traits, is probably a framework that conserves high sensitivity but also performs much better in terms of specificity.

Reassessing the Stoichiogenomic Theory in Multicellular Eukaryotes

This study suggests that stoichiogenomic N-saving mechanisms are absent from metazoan genomes. Elser et al. (2006) compared the proteomic N-cost of nine plants and nine animals and observed a lower average NARSC in plants (mean = 0.353) compared with animals (mean = 0.378). This observed difference of proteomic N-cost was interpreted as a stronger stoichiogenomic selection on N-savings in plants in response to environmentally limited N supplies. Indeed, plants extract inorganic N from their environment while animals obtain N through the amino acids contained in their diet. However, this plant–animal duality may not be so clear as the 26 isopod proteomes from this study have a NARSC ranging from 0.361 to 0.371 (mean = 0.367), that is, an intermediate value between plants and animals proteomes reported in Elser's study. So this apparent opposition between plants and metazoans may not hold when one increases the phylogenetic sampling. We believe that considering the small number of taxa, the absence of phylogenetic control, and the nonconsideration of alternative nonadaptive forces in most stoichiogenomic studies published so far, a global revision of the stoichiogenomic evidence in

Downloaded from http://mbe.oxfordjournals.org/ at INIST-CNRS on September 29, 2016

multicellular eukaryotes is warranted. Indeed, while stoichiogenomic mechanisms have been reliably evidenced in microorganisms, these latter also exhibit much larger effective population size than multicellular eukaryotes (Lynch and Conery 2003). Natural selection might not be efficient enough to retain weakly beneficial N-saving point mutations in multicellular eukaryotes.

Thinking Beyond Stoichiogenomics: Other Levels of Adaptive Responses

The fact that strong N limitation did not produce any composition change even in the most expressed proteins raises questions as to how organisms coped with this limitation and if efficient N-saving mechanisms could actually be selected either at specific genes or at higher organization levels. Indeed, previous studies evidenced that these compositional changes may be selected not in all or most-expressed genes but more specifically in the small pool of genes involved in N metabolism (Baudouin-Cornu et al. 2001) or expressed in response to N limitation (Gilbert and Fagan 2011; Grzymski and Dussaq 2012). Thus the composition of these particular transcripts and corresponding proteins (e.g., glutamine synthetase or leucine aminopeptidase) would be worth investigating, provided that functional annotations can be propagated based on arthropod reference genomes.

GW organisms evolved specific biological traits when colonizing N-depleted subterranean habitats, for example improved food-finding abilities and decreased metabolic rate (Hüppop 1987, 2005; Hervant and Renault 2002). Some GW isopods also display trophic specialization on sedimentary biofilm which is generally the most abundant food source in their habitats, resulting in optimized nutrient assimilation on this preferred food source (Francois et al. 2016). Because driven by a limited set of genes, morphological, behavioral, and physiological adaptations may also be much faster to set up than compositional changes, which require substitutions to accumulate over the whole transcriptome. In turn these adaptations at higher organization levels may have sufficiently reduced nutrient needs and/or improved nutrient acquisition for GW organisms not to require any other nutrient-saving mechanisms. This interplay between the evolution of different life-history traits has already been raised by Bragg and Wagner (2007) who did not detect the expected stoichiogenomic signal in the genome of yeast selected under carbon limitation.

In conclusion, the transcriptomes and proteomes of isopods evolving in N-limited habitats do not display any stoichiogenomic N-saving compositional changes. Some stoichiogenomic-like patterns were observed at the transcriptomic level but were far more compatible with nonadaptive processes. Altogether, this study casts doubt on the validity of the stoichiogenomic theory in metazoans and calls for a better consideration of the influence of nonadaptive processes. However, recent literature provided many evidence of stoichiogenomic signal in microorganisms. This discrepancy may be linked to differences in effective population sizes, natural selection being not efficient enough in metazoan species to select for fine-scale stoichiogenomic mechanisms. Nutrient-saving mechanisms in metazoans facing strong environmental limitations may also be more quickly selected at other biological levels (e.g., decrease of metabolic rate, increase of assimilation efficiency), these adaptations being sufficient to make further stoichiogenomic mechanisms unnecessary.

Materials and Methods

Definition and Sampling of the 13 Species Pairs

Using a large phylogeny of Asellidae (Morvan et al. 2013), we delimited 13 independent species pairs, where one species is subterranean and the other is a surface species (fig. 1). For each of the 26 selected species, individuals were sampled for RNA extraction (supplementary table S1, Supplementary Material online) and for elemental composition determination (see below), and flash frozen alive in the field.

Ecological Characterization

The estimation of the two ecological parameters (N availability and C:N mismatch) was achievable in 18 sampling sites corresponding to nine SW and nine GW habitats (table 1). At each of these sites, the sampling consisted in collecting on a depth of 1 cm all available trophic resources occurring at the surface of a standardized box (256 cm^2) (following the approach of Huntsman et al. 2011). This sampling was replicated three times in each site and was restricted to the trophic resources which are known to be edible by Asellidae. According to the literature, these trophic resources correspond to coarse particulate organic matter (1 mm $<$ particle size $<$ 6 mm), fine particulate organic matter (particle size $<$ 1 mm), roots, algae, mosses, and sedimentary biofilm developed on sand particles (200 μm $<$ particle size $<$ 1000 μm) (Moore 1975; Simon et al. 2003; Leberfinger et al. 2011; Mondy et al. 2014; Francois et al. 2016). Samples were flash frozen in the field. At the laboratory, all collected trophic resources were unfrozen and sorted according to the broad trophic categories defined previously. Trophic resources were then frozen, freeze-dried, and weighted to obtain a dry mass of each resource per square meter (g of Dry Weight (DW) . m^{-2}). Sedimentary biofilm samples were treated with 1N HCl to remove inorganic carbon (mainly carbonates) using the "capsule method," as described in Brodie et al. (2011). Elemental composition of each trophic resource (% C and % N, i.e., the organic carbon and nitrogen content of each resource as % of dry mass) was then measured with an elemental analyzer (Thermo FlashEA 1112, ThermoElectro, Milan, Italy). At each site, the abundance of all available resources (g of DW . m^{-2}) and their elemental composition (% N) were used to calculate the normalized mass of N (grams of N per square meter) contained in the trophic resources available for each species in its habitat, corresponding to the N availability (table 1). The coefficient of variation for N availability (based on three replicates) was lower than 15% at all sampling sites. Similarly, the C availability (grams of C . m^{-2}) was calculated and was used to estimate the average carbon-to-nitrogen ratio (C:N) of the bulk trophic resources available for each species in its habitat (called $C{:}N_{resources}$). Between 9 and 15 isopods were collected at each sampling site (except

Downloaded from http://mbe.oxfordjournals.org/ at INIST-CNRS on September 29, 2016

for *Proasellus parvulus* where only four individuals were available) and freeze-dried at the laboratory. Their elemental composition was then measured to determine the average C:N of each species (called $C{:}N_{isopods}$). For each species, the C:N mismatch between the available trophic resources and the isopods was then calculated following Elser and Hassett (1994):

$$C{:}N\ mismatch = C{:}N_{resources} - C{:}N_{isopods}$$

According to ecological stoichiometry theory (Sterner and Elser 2002), high C:N mismatch indicates strong N limitation in terms of trophic resources quality for the isopods.

Transcriptome Sequencing, Assembly, and Expression Estimates

Total RNA was isolated using TRI Reagent (Molecular Research Center). Extraction quality was checked on a BioAnalyser RNA chip (Agilent Technologies) and RNA concentrations were estimated using a Qubit fluorometer (Life Technologies). Prior to any additional analysis, species identification was corroborated for each individual by sequencing a fragment of 16S gene. Equimolar pools of at least five individuals were made to achieve 10 µg of RNA (supplementary table S1, Supplementary Material online). Volumes were reduced using a Concentrator-Plus (Eppendorf) to achieve approximately 10 µl. Double strand polyA-enriched cDNA were then produced using the Mint2 kit (Evrogen) following the standard protocol except for the first-strand cDNA synthesis, where the CDS-1 adapter was used with the plugOligo-Adapter of the Mint1 kit (5′_AAGCAGTGGTATCA ACGCAGAGTACGGGGG_P_3′). After sonication with a Bioruptor Nextgen UCD300 (Diagenode) and purification with MinElute (Qiagen), Illumina libraries were prepared using the NEBNext kit (New England BioLabs) and amplified using 22 unique indexed primers. After purification with MinElute, 400–500 bp fragments were size selected on an agarose gel. Libraries were paired-end sequenced on a HiSeq2000 sequencer (Illumina) using 100 cycles at the National High-throughput DNA Sequencing Center (Copenhagen, Denmark). Four libraries per lane were multiplexed, resulting in around 50 million reads per species (supplementary table S1, Supplementary Material online). Reads have been deposited to the European Nucleotide Archive and are available under the study ID PRJEB14193.

Adapters were clipped from the sequences, low quality read ends were trimmed (phred score < 30) and low quality reads were discarded (mean phred score < 25 or if remaining length < 19 bp) using fastq-mcf of the ea-utils package (Aronetsy 2013). Transcriptomes were denovo assembled using Trinity (version 2013-02-25, Grabherr et al. 2011). Open reading frames (ORFs) were identified with TransDecoder (http://transdecoder.sourceforge.net/, last accessed on Jul 6, 2016). For each assembled component, only the longest ORF was retained. Transcriptome assemblies have been deposited to the ENA (ID PRJEB14193). Expression of each contig was estimated using RSEM which used bowtie to map the reads back on the assembled contigs and then estimated the abundance of each transcript as transcripts per million transcripts (TPM; Li et al. 2010; Wagner et al. 2012). Expression data have been deposited to Zenodo (http://dx.doi.org/10.5281/zenodo.54584). Protein abundances were estimated from the corresponding mRNA abundances. Although this estimation ignores the influence of variations of rates of translation and protein degradation, recent studies demonstrated a very good correlation between mRNA and protein abundances (e.g., Csardi et al. 2015; Li and Biggin 2015).

Total N-Budget per Million Reads

The overall mRNA N budget was calculated per million transcripts using the RNAT defined as $RNAT = \sum_i (N_i * TPM_i)$, where N_i is the total number of N atoms being used in the ORF of gene i, and TPM_i is the number of transcripts i per million transcripts. By in silico translating the CDSs, we also estimated the RNAP defined as $RNAP = \sum_i (N_i * TPM_i)$, where N_i is this time the total number of N atoms used in the protein i. It should be noted that in this formula, protein abundances are approximated by their respective mRNA abundances. Although RNAP is not a direct measure of the amino acid N-usage of a cell, it is expected to be well correlated to it (Csardi et al. 2015; Li and Biggin 2015). RNAT and RNAP were estimated using all transcripts identified as ORFs by Transdecoder (number of ORFs per species ranging from 14,198 to 36,967; supplementary table S3, Supplementary Material online).

Average N-Cost per Site

The average N-cost per amino acid was calculated using the number of N Atoms per Residue Side-Chain (NARSC) defined as $NARSC = \sum_i (N_i)/\sum_i (l_i)$, where N_i is this time the total number of N atoms in the side chain of the amino acids of the protein i and l_i is the length of this protein. NARSC ranges from 0 to 3 (1 N atom for asparagine, glutamine, and lysine; 2 for histidine and tryptophan; 3 for arginine). Following the same logic, for CDS, we estimated the average number of N atoms per base in third position of the codon (NAB3). NAB3 ranges from 2 to 5 (2 N atoms for T/U, 3 N atoms for C and 5 N atoms for G and A). NARSC and NAB3 were estimated using all transcripts identified as ORFs by Transdecoder. Mitochondrial sequences were identified by BLAST (Altschul et al. 1990) and removed from the data set.

Families of Orthologous Genes

In addition to ORF as identified by Transdecoder, we also used BLASTx (Altschul et al. 1990) against complete proteomes from 20 representative arthropods species from the Ensembl Metazoa database to recover additional CDS. For each Trinity component we only kept one CDS, either the one showing the best BLASTx hit, or if no BLASTx match were found, the Transdecoder ORF that was the most expressed. Component showing no evidence of ORF have been discarded. For each CDS showing BLASTx hits, the best hit was selected to annotate start and stop position as well as frameshifts, by using Genewise (Birney et al. 2004). All sequences and their annotations were then loaded in an ACNUC database (Gouy et al. 1985). Gene families were

Downloaded from http://mbe.oxfordjournals.org/ at INIST-CNRS on September 29, 2016

then delimited using an all-against-all BLASTP (Altschul et al. 1990) and SiLix (Miele et al. 2011) including the 20 reference arthropod proteomes in the analysis. A first set of protein alignments and phylogenetic trees for each family was build and loaded in a gene family database (using the procedure described in Penel et al. [2009]). We then used a tree pattern matching algorithm (Dufayard et al. 2005) to search in all gene families for clades containing between 13 and 26 Asellidae sequences but no other arthropod sequences. That way large multigenic families are split in multiple orthologous sets of Asellidae sequences. After removal of any tree pattern that contained multiple sequences per species, we obtained 6,255 patterns hereafter considered as one-to-one orthologous genes. The CDS of each of these genes were aligned with Prank (Loytynoja and Goldman 2005) using a codon model, and Gblocks was used to select conserved positions (Castresana 2000; codon model and "half allowed gap positions" option). The 6,255 alignments have been deposited to Zenodo (http://dx.doi.org/10.5281/zenodo.53830).

Substitution Reconstruction and Associated N-Cost

For each pair of SW/GW species we defined two closely related species that were later used as outgroups (supplementary table S3, Supplementary Material online). For each gene, whenever possible, we defined sets of four sequences composed of a species pair and its two outgroups, hereafter referred as quartets (number of quartets per pair: min = 1,796; max = 3,229; supplementary table S3, Supplementary Material online). We labeled quartets as HE if the sequences of both species of the pair were within their 20%-most expressed transcripts (number of HE genes per pair: min = 311; max = 1,177; supplementary table S3, Supplementary Material online). For each quartet, all subsequent analyses considered only the sites (amino acid or third codon positions) which did not contain gaps and for which the outgroups and at least one species of the pair were identical, including sites where no difference was observed. The differences observed between the two focal species were then polarized using the parsimony criteria: the two outgroups defined the ancestral state and if any species of the pair displayed a different character state, it was then considered a derived state, that is, a substitution that happened on the terminal branch leading to this species. Parsimony tends to incorrectly infer the substitution history when homoplasies are common in a context of biased base composition (Eyre-Walker 1998). In our framework, parsimony will be especially sensitive to convergent substitutions that happen in the outgroup ancestor and in one of the extant species of a pair. The probability of such event will increase with the evolutionary distance between the focal species and their outgroups. The two pairs that are expected to be the most sensitive to this bias (p3 and p10) did not display any peculiar behavior compared with the other pairs (tables 3 and 4 and fig. 2). Thus, parsimony polarization errors did not seem to affect our capacity to test the stoichiogenomic theory.

We restricted this analysis to the inferred substitutions which changed the N-cost per site. For example, the substitutions A→G do not imply any variation of N-cost and thus were excluded from the analysis, whereas the substitutions A→T or C→G, which imply, respectively, a decrease (−3 N atoms) or an increase (+ 2 N atoms) of the N-cost, were considered in this analysis. The number of such N-cost-changing substitutions ranged from 643 to 6,322 for amino acids and from 2,965 to 46,161 for third codon positions (supplementary table S3, Supplementary Material online). Relative to SW species, GW species display life-history traits (e.g., higher longevity, lower metabolic rate, and longer generation time) often associated with lower substitution rate. Comparing directly the number of substitutions of the SW and GW species (e.g., the resultant change in N-cost across all substitutions) may thus be biased if far more substitutions occurred in the SW than in the GW branch since their last ancestor. Rather, we computed the proportion of N-costly substitutions which is normalized by the number of substitutions which have occurred in each branch. These proportions of N-costly substitutions in proteins and in third codon positions since the last common ancestor of each species pair (p_{AA} and p_{B3}, respectively) were calculated as follows:

$$p = (\text{number of inferred substitutions increasing the N-cost})/(\text{number of inferred substitutions changing the N-cost})$$

For each species, both proportions of N-costly substitutions were calculated globally on the concatenation of orthologous genes and proteins. We also used a bootstrap procedure (detailed in the following paragraph) to test the stoichiogenomic theory within each species pair.

On the exact same data set as for p_{B3}, we also calculated the GC3 of GW and SW species in each considered quartet. Within each pair, the GC3 of GW and SW species were compared across genes families with paired Wilcoxon tests.

Testing for Strand Asymmetries in Substitution Rates

When mutation and selection forces are strand-independent, complementary bases are present at the same frequency on a given DNA strand (A = T and C = G) and complementary substitutions occur at the same rate (e.g., $qC{\rightarrow}T = qG{\rightarrow}A$). In bacteria, during DNA replication, the lagging strand has a higher mutation rate, which generates an asymmetric substitution pattern between the two strands, the so-called GC-skew (Lobry 1996). Transcription also generates strand-asymmetric mutations by increasing the deamination of the coding strand, which generates a disequilibrium between the two complementary transitions C→T and G→A resulting in $qC{\rightarrow}T > qG{\rightarrow}A$ (Francino and Ochman 2001; Mugal et al. 2010). Under the stoichiogenomic theory for N-savings, one would expect selection to favor alleles reducing the N-cost of RNAs which would generate an asymmetric substitution pattern in CDS. Indeed, a transversion on the coding strand from A (costing five N atoms) to T (two N atoms) represents an economy of three N atoms for this position on the corresponding RNA. Likewise, G→C represents an economy of two N atoms, while the complementary T→A and C→G represent a burden of three and two more N atoms, respectively (fig. 4). It should be noted that such transversions affect the N-cost of RNAs, but not of DNA: a T→A (or C→G)

Downloaded from http://mbe.oxfordjournals.org/ at INIST-CNRS on September 29, 2016

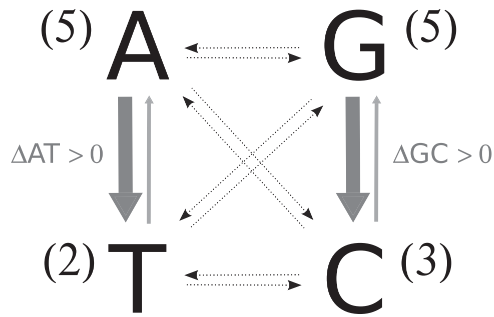

Fig. 4. Asymmetries in substitution rates under the N-saving stoichiogenomic theory. The substitutions we are interested in (i.e., those sensitive to stoichiogenomic selection but insensitive to gBGC and deamination) are represented by plain arrows. Numbers in brackets indicate the N-cost of each nucleotide (as number of N atoms). Arrow width is proportional to the expected substitution rate. $\Delta AT = (qA \rightarrow T)-(qT \rightarrow A)$ and $\Delta GC = (qG \rightarrow C)-(qC \rightarrow G)$, where $qX \rightarrow Y$ is the substitution rate from X to Y.

transversion on one DNA strand results in a A→T (G→C) transversion on the other strand, and hence remains neutral in terms of N-cost at the DNA level. Thus the stoichiogenomic theory predicts that selection should favor A→T and G→C substitutions specifically on the coding strand. We therefore used these two classes of complementary transversions to build a test of the stoichiogenomic theory for N-savings in transcriptomes (see fig. 4). The substitutional strand asymmetries are measured on the coding strand by:

$$\Delta AT = (qA \rightarrow T)-(qT \rightarrow A)$$

$$\Delta GC = (qG \rightarrow C)-(qC \rightarrow G)$$

Under the stoichiogenomic theory, ΔAT and ΔGC are expected to correlate with N limitation (i.e., $\Delta AT_{\text{N-limited species}} > \Delta AT_{\text{reference species}}$ and likewise for ΔGC). This stoichiogenomic test presents three advantages: first, it is not affected by the strand-asymmetric deamination process; second, it is robust to the potential impact of gBGC (which should affect exclusively A:T→G:C and G:C→A:T changes); and third, it can easily be applied in a comparative framework.

This test was performed on the orthologous genes quartets defined above. First we considered all substitutions, whether they changed the N-cost per site or not. The number of such substitutions at third codon positions ranged from 4,172 to 66,341 (supplementary table S3, Supplementary Material online). Substitution rate is defined as the rate at which sites switched from base X to base Y since the last common ancestor of a species pair. This rate takes into account the composition of the ancestral transcriptome and is calculated on third codon positions as follows:

$$qX \rightarrow Y = (\text{number of changes } X \rightarrow Y) / (\text{number of bases X in the ancestral transcriptome})$$

For each species, we then calculated the two measures of strand asymmetry, ΔAT and ΔGC, globally on the concatenation of all orthologous genes considered for its pair (i.e., quartets). We also used a bootstrap procedure to test within each pair if the stoichiogenomic theory was verified. The polarization of these unilateral within-pair tests (i.e., which species is the most N-limited in the pair) was defined in three different ways: 1) according to the habitat type (GW more N-limited than SW; habitat hypothesis); 2) according to the N limitation in quantity (N availability; quantity hypothesis); and 3) according to the N limitation in quality (C:N mismatch; quality hypothesis) (supplementary table S2, Supplementary Material online). To explain this bootstrap procedure, let us take the example of the habitat hypothesis, in which we expect that $\Delta AT_{GW} > \Delta AT_{SW}$ (GW more N-limited than SW), meaning that $\Delta AT_{GW} - \Delta AT_{SW} > 0$, the same applies to ΔGC. For each pair, this bootstrap procedure was iterated 1,000 times, and consisted in sampling with replacement N genes at each iteration (N being the total number of genes considered for this pair, that is, the number of quartets; supplementary table S3, Supplementary Material online). At each iteration, we calculated ΔAT_{GW}, ΔAT_{SW}, ΔGC_{GW},and ΔGC_{SW} on the concatenation of these N sampled gene families. The statistics we were interested in are the differences $\Delta 1 = \Delta AT_{GW} - \Delta AT_{SW}$ and $\Delta 2 = \Delta GC_{GW} - \Delta GC_{SW}$, which are expected to be significantly superior to 0 under the stoichiogenomic theory. The distributions of Δ1 and Δ2 were then centered around 0 by subtraction of their respective mean to create the "null" distributions $\Delta 1_{null}$ and $\Delta 2_{null}$ (i.e., corresponding to the null hypotheses $\Delta AT_{GW} = \Delta AT_{SW}$ and $\Delta GC_{GW} = \Delta GC_{SW}$). The global statistics $\Delta 1_{global}$ and $\Delta 2_{global}$ (calculated once on the concatenation of all gene families) can then be compared to these null distributions, by calculating the percentage of the centered statistics $\Delta 1_{null}$ and $\Delta 2_{null}$ which are strictly superior to $\Delta 1_{global}$ and $\Delta 2_{global}$, respectively. This percentage represents the *P* value of our unilateral bootstrap tests (null hypotheses $\Delta AT_{GW} = \Delta AT_{SW}$ and $\Delta GC_{GW} = \Delta GC_{SW}$; alternative hypotheses $\Delta AT_{GW} > \Delta AT_{SW}$ and $\Delta GC_{GW} > \Delta GC_{SW}$). The same principle applies for the two other ways to consider N limitation (quantity and quality hypotheses), the polarization of the test depending on the ecological parameters in each pair (N availability and C:N mismatch, respectively; see supplementary table S2, Supplementary Material online).

Testing for Changes in Transcript and Protein Expression

For each pair, this analysis was performed on the orthologous genes shared by both species of the pair. Orthologs displaying more than 15% of difference in N-cost between SW and GW species were discarded. The analysis was subsequently restricted to the 10% most N-costly gene families, either in terms of NARSC (proteins) or NAB3 (transcripts) (supplementary table S8, Supplementary Material online). The expression of these N-costly proteins/transcripts was estimated by their expression class (ten classes corresponding to the within-species deciles) and compared between SW and GW species with unilateral paired Wilcoxon tests. A decrease of the expression of these N-costly gene families is expected in the most N-limited species of each pair. As detailed in the previous paragraph, these within-pair tests were polarized in three

Downloaded from http://mbe.oxfordjournals.org/ at INIST-CNRS on September 29, 2016

different ways: according to the habitat (GW more N-limited than SW), to the N limitation in quantity (N availability) and in quality (C:N mismatch) (see supplementary table S2, Supplementary Material online).

Species Tree

Comparative analyses require an accurate representation of the phylogenetic relationships. We extracted a set of 386 1to1 orthologous gene alignments present in the 26 species. The concatenated alignment was used to reconstruct a phylogram with phyML with a general time reversible +G+I model of evolution (Guindon et al. 2010). Using this topology we estimated a chronogram with MCMCtree (Yang 2007) using 386 partitions, the formal likelihood function, independent rates model, and an Hasegawa–Kishino–Yano +G5 as model of substitutions. We used two independent runs of 42,000 iterations (10% as burn in) to check chains convergence and set the divergence between the *Bragasellus* and *Proasellus* to be not older than 150 Ma.

Statistical Analyses

Statistical analyses were performed with R 3.1 software (R Development Core Team 2014) and statistical significance was accepted at $P < 0.05$. The two ecological parameters (N availability and C:N mismatch) have been log-transformed prior to statistical analyses to correct for nonnormality of the data. According to the recommendations of Warton and Hui (2011) for proportion data, the proportions of N-costly substitutions (p_{AA} and p_{B3}) have been logit-transformed prior to statistical analyses. Phylogenetic Generalized Least-Squares PGLS models were used (Martins and Hansen 1997) to test the effect of ecological parameters on dependent variables using a likelihood ratio test between the models with and without the given explanation variable. Analyses were performed in R using the ape (version 3.2, Paradis et al. 2004), phytools (version 0.4, Revell 2012), geiger (version 2.0, Harmon et al. 2008), and nlme (version 3.1) packages. R^2 for PGLS models were calculated using likelihood-ratio based pseudo-R-squared (Magee 1990; Nagelkerke 1991) using the MuMIn package (version 1.13). In the within-pair tests, *P* values were subsequently corrected for multiple hypothesis testing using the false discovery rate method (Benjamini and Hochberg 1995; Benjamini and Yekutieli 2001).

Supplementary Material

Supplementary figure S1 and supplementary tables S1–S8 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors thank the following collaborators and organizations for their precious help in collecting biological material: Alhama de Granada municipality, M. Isabel Carreira, T. Datry, D. Eme, Cueva de Valporquero, Junta de Andalucia (Hundidero–Gato), and Rafael Orozco. The authors also thank Ludovic Orlando, Andaine Seguin-Orlando, and the laboratory technicians at the Danish High-throughput DNA Sequencing Centre. This work was performed using the computing facilities of the CNRS/IN2P3 Computing Center and of the CC LBBE/PRABI (Lyon/Villeurbanne–France). This work was supported by grants from the Agence Nationale de la Recherche (ANR-08-JCJC-0120-01 DEEP) and the Centre National de la Recherche Scientifique (APEGE 2012 "Stoichiogenomics" No. 70632 and EC2CO 2013–2014 "Ecosphere Continentale et Cotiere"—"CoCoNutS" project). The authors thank Ross Vander Vorste (Department of Forest Resources & Environmental Conservation, Virginia Tech., USA) for English editing. The authors thank the editor and two anonymous reviewers for their valuable comments on an earlier version of this manuscript.

References

Acquisti C, Elser JJ, Kumar S. 2009. Ecological nitrogen limitation shapes the DNA composition of plant genomes. *Mol Biol Evol*. 26:953–956.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol*. 215:403–410.

Aronesty E. 2013. Comparison of sequencing utility programs. *Tobioij* 7:1–8.

Baudouin-Cornu P, Surdin-Kerjan Y, Marlière P, Thomas D. 2001. Molecular evolution of protein atomic composition. *Science* 293:297–300.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B Methodol*. 57:289–300.

Benjamini Y, Yekutieli D. 2001. The control of the false discovery rate in multiple testing under dependency. *Ann Stat*. 29:1165–1188.

Birney E, Clamp M, Durbin R. 2004. GeneWise and genomewise. *Genome Res*. 14:988–995.

Bragg JG, Wagner A. 2007. Protein carbon content evolves in response to carbon availability and may influence the fate of duplicated genes. *Proc R Soc Lond B Biol Sci*. 274:1063–1070.

Bragg JG, Wagner A. 2009. Protein material costs: single atoms can make an evolutionary difference. *Trends Genet*. 25:5–8.

Brodie CR, Leng MJ, Casford JS, Kendrick CP, Lloyd JM, Yongqiang Z, Bird MI. 2011. Evidence for bias in C and N concentrations and δ13C composition of terrestrial and aquatic organic materials due to pre-analysis acid preparation methods. *Chem Geol*. 282:67–83.

Carter CG, Houlihan DF. 2001. Protein synthesis. *Fish Physiol*. 20:31–75.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol*. 17:540–552.

Cross WF, Benstead JP, Rosemond AD, Bruce Wallace J. 2003. Consumer-resource stoichiometry in detritus-based streams. *Ecol Lett*. 6:721–732.

Csárdi G, Franks A, Choi DS, Airoldi EM, Drummond DA. 2015. Accounting for experimental noise reveals that mRNA levels, amplified by post-transcriptional processes, largely determine steady-state protein levels in yeast. *PLoS Genet*. 11:e1005206.

Cuhel RL, Taylor CD, Jannasch HW. 1981. Assimilatory sulfur metabolism in marine microorganisms: sulfur metabolism, protein synthesis, and growth of *Pseudomonas halodurans* and *Alteromonas luteo-violaceus* during unperturbed batch growth. *Arch Microbiol*. 130:8–13.

Downing JA, Osenberg CW, Sarnelle O. 1999. Meta-analysis of marine nutrient-enrichment experiments: variation in the magnitude of nutrient limitation. *Ecology* 80:1157–1167.

Dufayard JF, Duret L, Penel S, Gouy M, Rechenmann F, Perrière G. 2005. Tree pattern matching in phylogenetic trees: automatic search for orthologs or paralogs in homologous gene sequence databases. *Bioinformatics* 21:2596–2603.

Duret L, Galtier N. 2009. Biased gene conversion and the evolution of mammalian genomic landscapes. *Annu Rev Genomics Hum Genet*. 10:285–311.

Downloaded from http://mbe.oxfordjournals.org/ at INIST-CNRS on September 29, 2016

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

Elser JJ, Bracken MES, Cleland EE, Gruner DS, Harpole WS, Hillebrand H, Ngai JT, Seabloom EW, Shurin JB, Smith JE. 2007. Global analysis of nitrogen and phosphorus limitation of primary producers in freshwater, marine and terrestrial ecosystems. *Ecol Lett*. 10:1135–1142.

Elser JJ, Fagan WF, Subramanian S, Kumar S. 2006. Signatures of ecological resource availability in the animal and plant proteomes. *Mol Biol Evol*. 23:1946–1951.

Elser JJ, Fagan WF, Denno RF, Dobberfuhl DR, Folarin A, Huberty A, Interlandi S, Kilham SS, McCauley E, Schulz KL, et al. 2000. Nutritional constraints in terrestrial and freshwater food webs. *Nature* 408:578–580.

Elser JJ, Hasset RP. 1994. A stoichiometric analysis of the zooplankton–phytoplankton interaction in marine and freshwater ecosystems. *Nature* 370:211–213.

Eyre-Walker A. 1998. Problems with parsimony in sequences of biased base composition. *J Mol Evol*. 47:686–690.

Felsenstein J. 1985. Phylogenies and the comparative method. *Am Nat*. 125:1.

Flamholz A, Phillips R, Milo R, Kellogg D, Darwin C. 2014. The quantified cell. *Mol Biol Cell*. 25:3497–3500.

Francino MP, Ochman H. 2001. Deamination as the basis of strand-asymmetric evolution in transcribed *Escherichia coli* sequences. *Mol Biol Evol*. 18:1147–1150.

Francois C, Mermillod-Blondin F, Malard F, Fourel F, Lécuyer C, Douady CJ, Simon L. 2016. Trophic ecology of groundwater species reveals specialization in a low-productivity environment. *Funct Ecol*. 30:262–273.

Gibert J, Deharveng L. 2002. Subterranean ecosystems: a truncated functional biodiversity. *Bioscience* 52:473.

Gilbert JDJ, Acquisti C, Martinson HM, Elser JJ, Kumar S, Fagan WF. 2013. GRASP [genomic resource access for stoichioproteomics]: comparative explorations of the atomic content of 12 *Drosophila* proteomes. *BMC Genomics* 14:599.

Gilbert JDJ, Fagan WF. 2011. Contrasting mechanisms of proteomic nitrogen thrift in Prochlorococcus. *Mol Ecol*. 20:92–104.

Gouy M, Gautier C, Attimonelli M, Lanave C, di Paola G. 1985. ACNUC–a portable retrieval system for nucleic acid sequence databases: logical and physical designs and usage. *Comput Appl Biosci*. 1:167–172.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*. 29:644–652.

Grzymski JJ, Dussaq AM. 2012. The significance of nitrogen cost minimization in proteomes of marine microorganisms. *ISME J*. 6:71–80.

Guindon S, Dufayard JF, Lefort V, Anisimova M. 2010. New algorithms and methods to estimate maximum- likelihoods phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. 59:307–321.

Günther T, Lampei C, Schmid KJ. 2013. Mutational bias and gene conversion affect the intraspecific nitrogen stoichiometry of the *Arabidopsis thaliana* transcriptome. *Mol Biol Evol*. 30:561–568.

Harmon LJ, Weir JT, Brock CD, Glor RE, Challenger W. 2008. GEIGER: investigating evolutionary radiations. *Bioinformatics* 24:129–131.

Hervant F, Renault D. 2002. Long-term fasting and realimentation in hypogean and epigean isopods: a proposed adaptive strategy for groundwater organisms. *J Exp Biol*. 205:2079–2087.

Hessen DO, Jeyasingh PD, Neiman M, Weider LJ. 2010. Genome streamlining and the elemental costs of growth. *Trends Ecol Evol*. 25:75–80.

Huntsman BM, Venarsky MP, Benstead JP, Huryn AD. 2011. Effects of organic matter availability on the life history and production of a top vertebrate predator (Plethodontidae: *Gyrinophilus palleucus*) in two cave streams. *Freshw Biol*. 56:1746–1760.

Hüppop K. 1987. Food-finding ability in cave fish (*Astyanax fasciatus*). *Int J Speleol*. 16:59–66.

Hüppop K. 2005. Adaptation to low food. In: Encyclopedia of caves. 2nd ed. Culver D.C. and White W.B. Amsterdam, Netherlands: Elsevier. pp. 4–9.

Karl DM, Michaels AF, Bergman B, Capone D, Carpenter E, Letelier R, Lipschultz F, Paerl H, Sigman D, Stal L. 2002. Dinitrogen fixation in the world's ocean. *Biogeochemistry* 57/58:47–98.

Karp PD, Ouzounis C, Moore-Kochlacs C, Goldovsky L, Kaipa P, Ahrén D, Tsoka S, Darzentas N, Kunin V, López-Bigas N. 2005. Expansion of the Biocyc collection of pathway/genome databases to 160 genomes. *Nucleic Acids Res*. 33:6083–6089.

Lane N, Martin W. 2010. The energetics of genome complexity. *Nature* 467:929–934.

Larsen T, Ventura M, O'Brien DM, Magid J, Lomstein BA, Larsen J. 2011. Contrasting effects of nitrogen limitation and amino acid imbalance on carbon and nitrogen turnover in three species of Collembola. *Soil Biol Biochem*. 43:749–759.

Leberfinger K, Bohman I, Herrmann J. 2011. The importance of terrestrial resource subsidies for shredders in open-canopy streams revealed by stable isotope analysis. *Freshw Biol*. 56:470–480.

Li JJ, Biggin MD. 2015. Statistics requantitates the central dogma. *Science* 347:1066–1067.

Li B, Ruotti V, Stewart RM, Thomson JA, Dewey CN. 2010. RNA-Seq gene expression estimation with read mapping uncertainty. *Bioinformatics* 26:493–500.

Li N, Lv J, Niu DK. 2009. Low contents of carbon and nitrogen in highly abundant proteins: evidence of selection for the economy of atomic composition. *J Mol Evol*. 68:248–255.

Lobry JR. 1996. Asymmetric substitution patterns in the two DNA strands of bacteria. *Mol Biol Evol*. 13:660–665.

Löytynoja A, Goldman N. 2005. An algorithm for progressive multiple alignment of sequences with insertions. *Proc Natl Acad Sci U S A*. 102:10557–10562.

Luo H, Thompson LR, Stingl U, Hughes AL. 2015. Selection maintains low genomic GC content in marine SAR11 lineages. *Mol Biol Evol*. 32:2738–2748.

Lv J, Li N, Niu DK. 2008. Association between the availability of environmental resources and the atomic composition of organismal proteomes: evidence from Prochlorococcus strains living at different depths. *Biochem Biophys Res Commun*. 375:241–246.

Lynch M, Conery JS. 2003. The origins of genome complexity. *Science* 302:1401–1404.

Magee L. 1990. R2 measures based on Wald and likelihood ratio joint significance tests. *Am Stat*. 44:250–253.

Martins EP, Hansen TF. 1997. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. *Am Nat*. 149:646.

Mazel D, Marlière P. 1989. Adaptive eradication of methionine and cysteine from cyanobacterial light-harvesting proteins. *Nature* 341:245–248.

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

Mermillod-Blondin F, Simon L, Maazouzi C, Foulquier A, Delolme C, Marmonier P. 2015. Dynamics of dissolved organic carbon (DOC) through stormwater basins designed for groundwater recharge in urban area: assessment of retention efficiency. *Water Res*. 81:27–37.

Miele V, Penel S, Duret L. 2011. Ultra-fast sequence clustering from similarity networks with SiLiX. *BMC Bioinformatics* 12:116.

Mondy N, Grossi V, Cathalan E, Delbecque JP, Mermillod-Blondin F, Douady CJ. 2014. Sterols and steroids in a freshwater crustacean (*Proasellus meridianus*): hormonal response to nutritional input. *Invertebr Biol*. 133:99–107.

Moore JW. 1975. The role of algae in the diet of *Asellus aquaticus* L. and *Gammarus pulex* L. *J Anim Ecol* 719–730.

Moore CM, Mills MM, Arrigo KR, Berman-Frank I, Bopp L, Boyd PW, Galbraith ED, Geider RJ, Guieu C, Jaccard SL, et al. 2013. Processes and patterns of oceanic nutrient limitation. *Nat Geosci*. 6:701–710.

Downloaded from http://mbe.oxfordjournals.org/ at INIST-CNRS on September 29, 2016

Morvan C, Malard F, Paradis E, Lefébure T, Konecny-Dupré L, Douady CJ. 2013. Timetree of aselloidea reveals species diversification dynamics in groundwater. *Syst Biol*. 62:512–522.

Mugal CF, Wolf JBW, Von Grünberg HH, Ellegren H. 2010. Conservation of neutral substitution rate and substitutional asymmetries in mammalian genes. *Genome Biol Evol*. 2:19–28.

Nagelkerke NJD. 1991. A note on a general definition of the coefficient of determination. *Biometrika* 78:691–692.

Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20:289–290.

Pascal G, Médigue C, Danchin A. 2006. Persistent biases in the amino acid composition of prokaryotic proteins. *Bioessays* 28:726–738.

Penel S, Arigon AM, Dufayard JF, Sertier AS, Daubin V, Duret L, Gouy M, Perrière G. 2009. Databases of homologous gene families for comparative genomics. *BMC Bioinformatics* 10(Suppl 6):S3.

Poulson TL, Lavoie KH. 2000. The trophic basis of subsurface ecosystems. Chapter 12. In: Wilkens, Culver D, Humphreys WF, editors. Subterranean ecosystems. Goodall, D.W. (editor in chief): Ecosystems of the World 30. Amsterdam: Elsevier Science B.V. pp. 231–249.

R Development Core Team. 2014. R: a language and environment for statistical computing. Vienna, Austria: R foundation for statistical computing. [cited 2016 Jul 6]. Available from: http://www.R-project.org/.

Revell LJ. 2012. Phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol*. 3:217–223.

Simon KS, Benfield EF, Macko SA. 2003. Food web structure and the role of epilithic biofilms in cave streams. *Ecology* 84:2395–2406.

Sterner RW, Elser JJ. 2002. *Ecological stoichiometry: the biology of elements from molecules to the biosphere*. Princeton (NJ): Princeton University Press.

Venarsky MP, Huntsman BM, Huryn AD, Benstead JP, Kuhajda BR. 2014. Quantitative food web analysis supports the energy-limitation hypothesis in cave stream ecosystems. *Oecologia* 176:859–869.

Vitousek PM, Howarth RW. 1991. Nitrogen limitation on land and in the sea: how can it occur? *Biogeochemistry* 13:87–115.

Wagner A. 2005. Energy constraints on the evolution of gene expression. *Mol Biol Evol*. 22:1365–1374.

Wagner GP, Kin K, Lynch VJ. 2012. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theor Biosci*. 131:281–285.

Warton DI, Hui FKC. 2011. The arcsine is asinine: the analysis of proportions in ecology. *Ecology* 92:3–10.

Webster JR, Meyer JL. 1997. Organic matter budgets for streams: a synthesis. *J N Am Benthol Soc* 16:141–161.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 24:1586–1591.

Downloaded from http://mbe.oxfordjournals.org/ at INIST-CNRS on September 29, 2016