

Why this genus? Mapping genus–domain relationships in microalgae research using bibliometrics and LLMs

Supplementary materials

Victor Pozzobon 

LGPM, CentraleSupélec, Université Paris-Saclay, Centre Européen de Biotechnologie et de Bioéconomie (CEBB), 3 rue des Rouges Terres 51110 Pomacle, France

1. Individual LLM output on the test cases

Identifying the best-suited LLM for the intended analysis is key. Indeed, even though prompt engineering can orient the output of a model, no amount of it will make up for an unfit reasoning capability. Therefore, several models were tested over a contrasted set of genera (*Chlorella* - 9771 articles, 145 items in the semantic network -, *Dunaliella* - 1451 articles, 62 items in the semantic network -, and *Galdieria* - 128 articles, 21 items in the semantic network -), relatively well-known to the author. In addition to pure performance analysis (relevance of the proposed concept), the difference between the three sets allows to evaluate the fitness of their context length and the sensitivity of the models to context dilution.

1 1. Qwen3 8B

Tables 1, 2, and 3 present the output for Qwen3 8B model. The first comments are technical. The model exhibits the expected behavior as temperature increases (more diverse output, higher number of concepts reported, even though sometimes lumped). This is particularly true for *Chlorella* genus. Still, the context tends to saturate (two missing trials for *Chlorella* and *Galdieria* genera and one for *Dunaliella* genus, confirmed with a rerun). Moving to the content, the model tends to identify only the main ones (e.g., biofuel and wastewater treatment for *Chlorella* genus) with a certain level of redundancy. In addition, it fails to pinpoint β -carotene for *Dunaliella* genus and phycocyanin for *Galdieria* genus. Overall, the model is not suited for the intended use, but shows some merit, especially given its low number of parameters.

1 2. Qwen3 30B A3B Instruct 2507

Qwen3 30B A3B Instruct 2507 offered minimal output with a tendency to lump concepts (Tables 4, 5, and 6), reducing *Chlorella* genus to only five entries, *Dunaliella* and *Galdieria* to two. Furthermore, its sensitivity to temperature is almost non-existent. In an attempt to improve the output fitness for the task at hand, the temperature value was increased up to 2, but it did not modulate the outcome significantly. Therefore, it can be deemed that this model reduces the complexity of the network too much. Even though prompt engineering could help manage concept lumping, the lack of sensitivity

to temperature restricts the possibilities offered by this tool. Finally, it should be noted that, while failing to identify the relevance of phycocyanin for *Galdieria* genus, it successfully identified carotenoids for *Dunaliella* genus.

1 3. Qwen3 30B A3B Thinking 2507

Qwen3 30B A3B Thinking 2507 suffered the same overreduction problem as Qwen3 30B A3B Instruct 2507, but to a lower extent (Tables 7, 8, and 9). Yet, it showed a much greater diversity in the naming of the concepts. Furthermore, it was able to identify the relevance of β -carotene for *Dunaliella* genus, twice over 6 runs, and phycocyanin for *Galdieria* genus. Still, while more promising than Qwen3 30B A3B Instruct 2507, in addition to sometimes saturating the context (twice for *Chlorella* genus), the results remain of too low quality to hope that even extensive prompt engineering could allow this model to perform the intended task properly.

1 4. Qwen3.5 27B

Qwen3.5 27B (Tables 10, 11, and 12) exhibits a clear tendency to agglomerate concepts to present them as pairs. Still, they are most of the time relevant. Furthermore, the model exhibits adequate sensitivity to the selected temperature level (slightly reordering concepts, modifying wording, but not altering overall coherence). Yet, it was not able to identify carotenoids specifically for *Dunaliella* genus (referring to *metabolites* instead), but successfully spotted phycocyanin for *Galdieria* genus. Overall, the model appears adequate for the task at hand, after a prompt fine-tuning analysis.

1 5. Gemma 2 27B It

Tables 13, 14, and 15 present the output for Gemma 2 27B It model. The first comments are technical, indeed, despite a maximum context length of 8192 tokens, the model is able to manage the relatively long prompt associated with *Chlorella* genus. Then, temperature has a very limited effect on the model output in the tested range (0 to 0.1). Indeed, neither the number of identified concepts nor their rankings is affected by this parameter. Focusing on the content, while they appear in a reasonable amount and show limited overlap for *Chlorella* genus, they are plenty and redundant for

Dunaliella genus. Finally, detailed identification does not seem within reach of the model as it barely mentions β -carotene for *Dunaliella* genus, not phycocyanin for *Galdieria* genus. All in all, while showing merit, this model is not suited for the task at hand.

1 6. Gemma 4 31B

Tables 16, 17, and 18 present the output for Gemma 4 31B model. Like its former iteration, this model exhibits low sensitivity to temperature in the tested range (0 to 0.1). Yet, the results can be deemed superior to those of Gemma 2 27B. It, as this iteration identifies the main concepts for *Chlorella* and *Dunaliella* genera successfully, while missing phycocyanin for *Galdieria* genus. Finally, this model exhibits a clear propensity to lump concepts. All in all, it is only this last point that advocates against the use of this model.

1 7. Llama 3.3 70B Instruct

Tables 19, 20, and 21 present the output for Llama 3.3 70B Instruct model. The concepts identified by the model are systematically relevant and well defined for *Chlorella* and *Dunaliella* genera, while they are a bit too numerous for *Galdieria* genus when the temperature is set to 0.1. Still, for all of the tested genera, it identified the primary and secondary concepts, such as photosynthesis for *Chlorella* genus, or aquaculture feed for *Dunaliella* genus. In addition, over all the trials, it identified β -carotene for *Dunaliella* genus and phycocyanin for *Galdieria* genus. The model can therefore be deemed fit for the intended use with minimal prompt engineering.

1 8. GPT OSS 120B

Tables 22, 23, and 24 present the output for GPT OSS 120B model. Starting with the technical aspects, the model shows a relatively high sensitivity to temperature, with an average consistency between the trials performed at a temperature of 0.10. In addition, the model tends to lump some concepts or use verbose formulations to describe them. Moving the results themselves, most of the time, the model identifies relevant concepts. Despite the variety in its responses, the model fails to mention β -carotene for *Dunaliella* genus. On the contrary, it systematically reports phycocyanin for *Galdieria* genus. Overall, the model can be deemed fit for the task at hand, but would require extensive prompt engineering, specifically oriented towards model guidance, as the regular prompt is already quite directive.

1 9. Mixtral 8x7B Instruct & Phi 3.5 MoE Instruct

Mixtral 8x7B Instruct and Phi 3.5 MoE Instruct results are not reported. Indeed, they consisted of a restitution of the most prominent keywords (about 50 for *Chlorella* genus, 40 for *Dunaliella* genus, and 10 for *Galdieria* genus), making it unfit for the task at hand.

1 10. LLM preliminary screening results tables

Genus	Temperature	N _{trial}	Order	Output Concept	Score
<i>Chlorella</i>	0.0	1	1	Biomass Production for Biofuel	53129.0
			2	Wastewater Treatment	33362.0
			3	Process Optimization	38200.0
			4	Lipid Production for Biofuel	28281.0
			5	Photosynthesis Research	8212.0
	0.1	2	1	Biomass production for biofuel	95.0
			2	Lipid production for biofuel	90.0
			3	Wastewater treatment	85.0
			4	Photosynthesis research	80.0
			5	Biotechnology applications	75.0
		3	6	Nutrient removal in wastewater	70.0
			7	Carbon capture and sequestration	65.0
			8	Algae-based biomass and lipid production	60.0
			9	Microalgae cultivation	55.0
			10	Energy production from algae	50.0
		4	1	Lipid Production for Biofuel	20060.0
			2	Biomass Production	19350.0
			3	Bioproduct Development	17146.0
			4	Wastewater Treatment	14916.0
			5	Photosynthesis Research	8661.0
		5	1	Biomass production for biofuel and nutrient removal	55788.0
			2	Wastewater treatment and nutrient removal	33362.0
			3	Photosynthesis and lipid production in microalgae	108367.0
			4	Bioproduct development and potential applications	16338.0
			5	Bioprocess optimization and system performance	38200.0

Table 1. Qwen3 8B output for *Chlorella* genus

Genus	Temperature	N _{trial}	Order	Output Concept	Score
<i>Dunaliella</i>	0.0	1	1	Algal biomass and protein production	7523.0
			2	Algal species and strains	5777.0
			3	Algal cultivation conditions	5745.0
			4	Algal treatment processes	2987.0
			5	Algal lipid and carotenoid production	2658.0
	0.1	2	1	Microalgae production and cultivation	8398.0
			2	Lipid and protein production	6893.0
			3	Experimental effects and treatments	6534.0
			4	Algal species and strains	5288.0
			5	Growth conditions	5585.0
		3	1	Biomass production	6893.0
			2	Microalgae species	5288.0
			3	Cultivation conditions	4555.0
			4	Bioactive compound production	2528.0
		5	1	Bioproduct development	6893.0
			2	Algal species	5288.0
			3	Physiological factors	5595.0
			4	Bioactive compound production	2528.0
			5	Treatment effects	7045.0

Table 2. Qwen3 8B output for *Dunaliella* genus

Genus	Temperature	N _{trial}	Order	Output Concept	Score
<i>Galdieria</i>	0.0	1	1	Algal biomass production	nan
			2	Wastewater treatment for biomass and bioproducts	nan
	0.1	1	1	Microalgae production	481.0
			2	Wastewater treatment	451.0
		2	1	Algal Cultivation	320.0
			2	Biomass Production from Wastewater	132.0
		4	1	Microalgae production	2636.0
			2	Wastewater treatment	1065.0

Table 3. Qwen3 8B output for *Galdieria* genus

Genus	Temperature	N _{trial}	Order	Output Concept	Score
<i>Chlorella</i>	0.0	1	1	Microalgae cultivation and biomass production	98765
			2	Lipid production and biofuel development	72432
			3	Wastewater treatment and nutrient removal	68450
			4	Process optimization and system performance	58920
			5	Bioproducts and biotechnology applications	45678
		1	1	Microalgae cultivation and biomass production	98765
			2	Lipid production and biofuel development	72432
			3	Wastewater treatment and nutrient removal	68450
			4	Process optimization and system performance	58210
			5	Bioproducts and biotechnology applications	45678
		2	1	Microalgae cultivation and biomass production	98765
			2	Lipid production and biofuel development	78456
			3	Wastewater treatment and nutrient removal	67345
			4	Process optimization and system performance	56234
			5	Bioproducts and biotechnology applications	45123
	0.1	3	1	Microalgae cultivation and biomass production	98765
			2	Lipid production and biofuel development	72432
			3	Wastewater treatment and nutrient removal	68412
			4	Process optimization and system performance	58943
			5	Bioproducts and biotechnology applications	45678
		4	1	Microalgae cultivation and biomass production	9875
			2	Lipid production and biofuel development	7642
			3	Wastewater treatment and nutrient removal	6893
			4	Process optimization and system performance	5921
			5	Biochemical composition and bioproduct synthesis	4230
		5	1	Microalgae cultivation and biomass production	98765
			2	Lipid production and biofuel development	78456
			3	Wastewater treatment and nutrient removal	69843
			4	Process optimization and system performance	58321
			5	Bioproducts and biotechnology applications	45678

Table 4. Qwen3 30B A3B Instruct 2507 output for *Chlorella* genus

Genus	Temperature	N _{trial}	Order	Output Concept	Score
<i>Dunaliella</i>	0.0	1	1	Lipid production for biofuel	8398
			2	Microalgae cultivation and growth optimization	8398
	0.1	1	1	Lipid production for biofuel	8398
			2	Microalgae cultivation and growth optimization	8398
		2	1	Lipid production for biofuel	8398
			2	Microalgae cultivation and growth optimization	8398
		3	1	Lipid production for biofuel	8398
			2	Microalgae cultivation and growth optimization	8398
		4	1	Lipid production for biofuel	8398
			2	Microalgae cultivation and growth optimization	8398
		5	1	Lipid production for biofuel	8398
			2	Microalgae cultivation and growth optimization	8398

Table 5. Qwen3 30B A3B Instruct 2507 output for *Dunaliella* genus

Genus	Temperature	N _{trial}	Order	Output Concept	Score
<i>Galdieria</i>	0.0	1	1	Microalgae cultivation and biomass production	412
			2	Wastewater treatment and nutrient recovery	345
	0.1	1	1	Microalgae cultivation and biomass production	412
			2	Wastewater treatment and nutrient recovery	358
		2	1	Microalgae cultivation and biomass production	412
			2	Wastewater treatment and nutrient recovery	345
		3	1	Microalgae cultivation and biomass production	417
			2	Wastewater treatment and nutrient recovery	360
		4	1	Microalgae cultivation and biomass production	417
			2	Wastewater treatment and nutrient recovery	360
		5	1	Microalgae cultivation and biomass production	417
			2	Wastewater treatment and nutrient recovery	360

Table 6. Qwen3 30B A3B Instruct 2507 output for *Galdieria* genus

Genus	Temperature	N _{trial}	Order	Output Concept	Score
<i>Chlorella</i>	0.0	1	1	Wastewater treatment	52269
			2	Lipid production for biofuel	35106
			3	Biomass extraction and yield	24507
			4	Photosynthetic protein synthesis	16255
			5	Biofuel production	1739
		1	1	Wastewater treatment	71765
			2	Lipid production for biofuel	50808
			3	Microalgae biomass composition	37103
			4	Microalgae cultivation and processing	31188
			5	Biofuel production	2133
	0.1	2	1	Microalgae wastewater treatment and lipid production	50112
			2	Microalgae wastewater treatment and nutrient removal	33065
			3	Microalgae growth and biochemical composition	22643
			4	Microalgae biofuel production	16298
			5	Microalgae growth and protein production	10445
			6	Microalgae growth and protein production	10445
			7	Microalgae growth and lipid extraction	9741
		4	1	Wastewater treatment and nutrient removal	9543
			2	Lipid production for biofuel	8683
			3	Algal biochemical composition and photosynthesis	4653
		4	4	Biofuel production	2133
			5	Extraction process	734

Table 7. Qwen3 30B A3B Thinking 2507 output for *Chlorella* genus

Genus	Temperature	N _{trial}	Order	Output Concept	Score
<i>Dunaliella</i>	0.0	1	1	Microalgae growth conditions	2883
			2	Microalgae lipid and protein production	1448
			3	Microalgae biomass production	1286
			4	Marine microalgae species and strains	1262
			5	Treatment effects	317
	0.1	1	1	Microalgae culture conditions	2082
			2	Marine microalgae species and strains	999
			3	Microalgae biomass and bioproduct cultivation	922
			4	Microalgae biochemical composition	903
		2	1	Carotenoid, fatty acid, lipid, and protein production	966
			2	Salinity and nitrogen effects on growth	775
			3	Biomass and bioproduct cultivation	608
			4	Marine microalgae species	410
		3	1	Microalgae growth conditions	840
			2	Microalgae biochemical composition	779
			3	Microalgae bioproduct production	749
			4	Marine microalgae species and strains	634
		4	1	Microalgae cell growth	1389
			2	Biomass and bioproduct production	1238
			3	Carotenoid and lipid production	1090
			4	Microalgae strain selection	411
		5	1	Pigment and lipid production	933
			2	Specific microalgae species	199

Table 8. Qwen3 30B A3B Thinking 2507 output for *Dunaliella* genus

Genus	Temperature	N _{trial}	Order	Output Concept	Score
<i>Galdieria</i>	0.0	1	1	Phycocyanin production	310
			2	Wastewater-based biomass production	132
	0.1	1	1	Phycocyanin production from microalgae	198
			2	Wastewater biomass production	132
		2	1	Phycocyanin production	262
			2	Wastewater treatment	168
		3	1	Phycocyanin production	310
			2	Wastewater bioproduct production	132
		4	1	Phycocyanin production in microalgae	220
			2	Wastewater-based bioproduct production	110
		5	1	Microalgae cultivation for phycocyanin	220
			2	Biomass production from wastewater	110

Table 9. Qwen3 30B A3B Thinking 2507 output for *Galdieria* genus

Genus	Temperature	N _{trial}	Order	Output Concept	Score
<i>Chlorella</i>	0.0	1	1	Lipid and Biofuel Production	10272
			2	Wastewater Treatment and Nutrient Removal	7483
			3	Process Optimization and Extraction	4611
			4	Physiological and Biochemical Analysis	3499
			5	High-value Bioproducts	2133
	0.1	1	1	Lipid and Biofuel Production	267585
			2	Wastewater Treatment and Nutrient Removal	158626
			3	Process Optimization and Extraction	131116
			4	Algal Physiology and Pigment Production	64647
		2	1	Biomass and Lipid Production for Biofuel	42336
			2	Algal Physiology and Pigment Analysis	31426
			3	Process Optimization and Modeling	19989
			4	Wastewater Treatment and Nutrient Removal	17721
		3	1	Microalgae Cultivation and Biomass Production	14524
			2	Wastewater Treatment and Nutrient Removal	6537
			3	Process Optimization and Extraction	5276
			4	Microalgae Biochemistry and Physiology	3499
		4	1	Lipid production for biofuel	148845
			2	Wastewater treatment and nutrient removal	128910
			3	Process optimization and extraction	79003
			4	Microalgae physiology and pigment production	75460
			5	High-value bioproducts	43051
		5	1	Lipid and Biofuel Production	224003
			2	Wastewater Treatment and Nutrient Removal	163470
			3	Cultivation Optimization and Modeling	95788
			4	Biochemical Composition and Physiology	64647

Table 10. Qwen3.5 27B output for *Chlorella* genus

Genus	Temperature	N _{trial}	Order	Output Concept	Score
<i>Dunaliella</i>	0.0	1	1	Biomass and Bioproduct Production	13906
			2	Biochemical Composition Analysis	12461
			3	Cultivation Environmental Conditions	11504
			4	Algal Strain Selection	4480
	0.1	1	1	Biomass and bioproduct production	1584
			2	Biochemical composition analysis	1090
			3	Growth and environmental factors	1084
			4	Treatment effects and activity	580
			5	Algal species characterization	298
		2	1	Biochemical metabolite production	12461
			2	Environmental cultivation factors	9312
			3	Bioproduct applications	3346
			4	Microalgal strain selection	3207
		3	1	Microalgae biomass production	12896
			2	Biochemical composition analysis	12461
			3	Cultivation environmental factors	12173
			4	Microalgae strain selection	3207
		4	1	Biochemical Metabolite Production	1090
			2	Cultivation and Environmental Factors	853
			3	Physiological Activity and Treatment	455
			4	Microalgal Species Selection	298
		5	1	Biochemical composition and metabolite production	1090
			2	Cultivation conditions and optimization	853
			3	Microalgal species and strains	510

Table 11. Qwen3.5 27B output for *Dunaliella* genus

Genus	Temperature	N _{trial}	Order	Output Concept	Score
<i>Galdieria</i>	0.0	1	1	Phycocyanin production	2080
			2	Wastewater treatment	1133
		1	1	Phycocyanin production	1722
			2	Wastewater treatment and biomass	1133
	0.1	2	1	Cultivation and phycocyanin production	2080
			2	Wastewater treatment and biomass production	1133
		3	1	Phycocyanin production and cultivation	320
			2	Wastewater treatment and biomass	168
		4	1	Phycocyanin cultivation optimization	329
			2	Wastewater treatment and biomass	281
		5	1	Phycocyanin cultivation optimization	256
			2	Wastewater biomass production	168

Table 12. Qwen3.5 27B output for *Galdieria* genus

Genus	Temperature	N _{trial}	Order	Output Concept	Score
<i>Chlorella</i>	0.0	1	1	Lipid production	100
			2	Wastewater treatment	95
			3	Biomass production	90
			4	Microalgae cultivation	85
			5	Biofuel production	80
			6	Growth and productivity	75
			7	Nutrient removal	70
			8	Biodiesel production	65
			9	Photosynthesis	60
			10	Microalgae species	55
		1	1	Lipid production	100
			2	Wastewater treatment	95
			3	Biomass production	90
			4	Microalgae cultivation	85
			5	Biofuel production	80
			6	Growth and productivity	75
			7	Nutrient removal	70
			8	Biodiesel production	65
			9	Photosynthesis	60
			10	Microalgae species	55
	0.1	2	1	Lipid production	120
			2	Wastewater treatment	100
			3	Biomass production	95
			4	Microalgae cultivation	90
			5	Biofuel production	85
			6	Growth and productivity	80
			7	Nutrient removal	75
			8	Biodiesel production	70
			9	Photosynthesis	65
			10	Microalgae species	60
		3	1	Lipid production	100
			2	Wastewater treatment	95
			3	Biomass production	90
			4	Microalgae cultivation	85
			5	Biofuel production	80
			6	Growth and productivity	75
			7	Nutrient removal	70
			8	Biodiesel production	65
			9	Photosynthesis	60
			10	Algal biomass	55
		4	1	Wastewater treatment	107
			2	Biofuel production	98
			3	Biomass production	92
			4	Lipid production	85
			5	Microalgae cultivation	78
			6	Growth and Productivity	72
			7	Nutrient removal	68
			8	Biodiesel production	65
			9	Microalgae biomass	62
			10	Photosynthesis	58
		5	1	Wastewater treatment	100
			2	Biofuel production	95
			3	Biomass production	90
			4	Lipid production	85
			5	Microalgae cultivation	80
			6	Growth and productivity	75
			7	Nutrient removal	70
			8	Biodiesel production	65
			9	Microalgae species	60
			10	Photosynthesis	55

Table 13. Gemma 2 27B It output for *Chlorella* genus

Genus	Temperature	N _{trial}	Order	Output Concept	Score
<i>Dunaliella</i>	0.0	1	1	Microalgae Cultivation	8398
			2	Lipid Production	2217
			3	Growth Conditions	5585
			4	Biomass Production	3731
			5	Nutrient Effects	1884
			6	Species Identification	5288
			7	Bioproduct Applications	3731
			8	Environmental Factors	1833
			9	Physiological Responses	2482
			10	Experimental Design	2115
	0.1	1	1	Microalgae Cultivation	100
			2	Lipid Production	85
			3	Growth and Biomass	75
			4	Bioproduct Application	60
			5	Environmental Factors	55
		2	1	Microalgae Cultivation	8398
			2	Lipid Production	2217
			3	Growth Conditions	5585
			4	Biomass Production	3731
			5	Nutrient Effects	1884
		3	1	Microalgae Cultivation	8398
			2	Biomass Production	3731
			3	Growth Conditions	5585
			4	Lipid Production	2217
			5	Nutrient Effects	1694
			6	Species Identification	5288
			7	Bioproduct Applications	3731
			8	Physiological Response	2482
			9	Environmental Factors	1884
			10	Culture Techniques	3617
		4	1	Microalgae Cultivation	8398
			2	Lipid Production	2217
			3	Growth Conditions	5585
			4	Biomass Production	3731
			5	Nutrient Effects	1884
		5	1	Microalgae Cultivation	8398
			2	Lipid Production	2217
			3	Growth and Development	5585
			4	Biomass Production	3731
			5	Environmental Effects	5273
			6	Nutrient Analysis	1690
			7	Species Identification	5288
			8	Bioproduct Applications	3731
			9	Experimental Conditions	4555
			10	Treatment Effects	2192

Table 14. Gemma 2 27B It output for *Dunaliella* genus

Genus	Temperature	N _{trial}	Order	Output Concept	Score
<i>Galdieria</i>	0.0	1	1	Microalgae Cultivation	10
			2	Bioproduct Production	6
			3	Wastewater Treatment	4
	0.1	1	1	Microalgae Cultivation	60
			2	Bioproduct Production	35
			3	Wastewater Treatment	31
		2	1	Microalgae Cultivation	10
			2	Bioproduct Production	6
			3	Wastewater Treatment	4
		3	1	Microalgae Cultivation	10
			2	Bioproduct Production	6
			3	Wastewater Treatment	4
		4	1	Microalgae Cultivation	10
			2	Bioproduct Production	6
			3	Wastewater Treatment	4
		5	1	Microalgae Cultivation	10
			2	Bioproduct Production	6
			3	Wastewater Treatment	4

Table 15. Gemma 2 27B It output for *Galdieria* genus

Genus	Temperature	N _{trial}	Order	Output Concept	Score
<i>Chlorella</i>	0.0	1	1	Biomass and Lipid Production for Biofuel	115420
			2	Wastewater Treatment and Nutrient Removal	78210
			3	Microalgae Species and Physiological Analysis	65400
			4	Process Optimization and Yield Analysis	52100
			5	Bioproducts and Biofuel Development	31200
	1	1	1	Biomass and Lipid Production for Biofuel	115432
			2	Wastewater Treatment and Nutrient Removal	68412
			3	Process Optimization and Yield Analysis	52145
			4	Microalgae Biological Properties and Proteins	42108
			5	Biofuel and Bioproduct Development	21540
	2	2	1	Biomass and Lipid Production for Biofuel	115420
			2	Wastewater Treatment and Nutrient Removal	78540
			3	Process Optimization and Yield Analysis	62110
			4	Biofuel and Bioproduct Development	43151
			5	Microalgae Species and Photosynthesis	32100
	0.1	3	1	Biomass and Lipid Production for Biofuel	115420
			2	Wastewater Treatment and Nutrient Removal	78214
			3	Process Optimization and Yield Analysis	62145
			4	Biofuel and Bioproduct Development	43151
			5	Microalgae Physiological and Biochemical Analysis	38412
	4	4	1	Biomass and Lipid Production for Biofuel	115432
			2	Wastewater Treatment and Nutrient Removal	68412
			3	Process Optimization and Yield Analysis	52145
			4	Microalgae Biological Properties and Proteins	42108
			5	Bio-products and Biofuel Development	28120
	5	5	1	Biomass and Lipid Production for Biofuel	115420
			2	Wastewater Treatment and Nutrient Removal	78245
			3	Process Optimization and Yield Analysis	62140
			4	Bioproducts and Biofuel Development	43151
			5	Microalgae Physiological and Species Analysis	38410

Table 16. Gemma 4 31B output for *Chlorella* genus

Genus	Temperature	N _{trial}	Order	Output Concept	Score
<i>Dunaliella</i>	0.0	1	1	Microalgae biomass production	15540
			2	Growth and cultivation conditions	12450
			3	Carotenoid and protein synthesis	8500
			4	Microalgae species and strain analysis	4100
			5	Treatment activity and response	3000
		1	1	Microalgae biomass production	15450
			2	Protein and lipid synthesis	10341
			3	Growth kinetics and environmental response	8540
			4	Microalgae strain characterization	4125
			5	Treatment activity analysis	3110
		2	1	Microalgae biomass production	14500
			2	Bioactive compound and protein synthesis	8500
			3	Growth optimization and environmental response	7800
			4	Microalgae strain characterization	5200
			5	Treatment activity analysis	3100
	0.1	3	1	Microalgae biomass production	15450
			2	Protein and lipid synthesis	10300
			3	Growth optimization and kinetics	8500
			4	Species strain characterization	7500
			5	Treatment activity analysis	4000
		4	1	Microalgae biomass production	15514
			2	Bioactive compound and protein synthesis	11331
			3	Growth kinetics and environmental response	10850
			4	Microalgae strain characterization	8224
			5	Treatment activity and efficacy	4112
		5	1	Microalgae biomass production	15450
			2	Protein and lipid synthesis	10500
			3	Growth and cultivation conditions	8200
			4	Microalgae strain characterization	6100
			5	Treatment activity analysis	3200

Table 17. Gemma 4 31B output for *Dunaliella* genus

Genus	Temperature	N _{trial}	Order	Output Concept	Score
<i>Galdieria</i>	0.0	1	1	Microalgae cultivation and production	180
			2	Wastewater-based biomass production	110
	0.1	1	1	Microalgae cultivation and production	180
			2	Wastewater-based biomass production	110
		2	1	Microalgae cultivation and production	175
			2	Wastewater-based biomass production	114
		3	1	Microalgae cultivation and production	180
			2	Wastewater-based biomass production	110
		4	1	Microalgae cultivation and production	175
			2	Wastewater-based biomass production	114
		5	1	Microalgae cultivation and production	210
			2	Wastewater-based biomass production	115

Table 18. Gemma 4 31B output for *Galdieria* genus

Genus	Temperature	N _{trial}	Order	Output Concept	Score
<i>Chlorella</i>	0.0	1	1	Microalgae cultivation	6244
			2	Biomass production	2724
			3	Biofuel production	1325
			4	Wastewater treatment	1653
			5	Lipid production	1421
			6	Growth rate	3426
			7	Photosynthesis research	449
			8	Biotechnology	706
			9	Carbon capture	684
			10	Nutrient removal	998
		1	1	Microalgae cultivation	6244
			2	Biomass production	2724
			3	Biofuel production	1325
			4	Wastewater treatment	1653
			5	Lipid production	1421
			6	Photosynthesis research	449
			7	Biotechnology	706
			8	Carbon capture	684
			9	Nutrient removal	998
			10	Microalgal biomass	544
		2	1	Microalgae cultivation	6244
			2	Biomass production	2724
			3	Biofuel production	1325
			4	Wastewater treatment	1653
			5	Lipid production	1421
			6	Growth rate	3426
			7	Photosynthesis research	449
			8	Biotechnology	706
			9	Carbon capture	684
			10	Nutrient removal	998
	0.1	3	1	Lipid production for biofuel	28281
			2	Microalgae cultivation	108367
			3	Wastewater treatment	33362
			4	Biomass production	53129
			5	Biofuel production	26713
			6	Carbon capture and sequestration	26186
			7	Photosynthesis research	8212
			8	Algal biomass production	8822
			9	Biotechnology applications	13109
			10	Nutrient removal	20557
		4	1	Lipid production for biofuel	28281
			2	Microalgae cultivation	108367
			3	Wastewater treatment	33362
			4	Biomass production	53129
			5	Biofuel production	26713
			6	Algal biomass production	8822
			7	Carbon capture and sequestration	13557
			8	Photosynthesis research	8212
			9	Protein production	26739
			10	Biotechnology applications	13109
		5	1	Microalgae cultivation	6244
			2	Biomass production	2724
			3	Biofuel production	1325
			4	Wastewater treatment	1653
			5	Lipid production	1421
			6	Growth rate	3426
			7	Photosynthesis research	449
			8	Biotechnology	706
			9	Carbon capture	684
			10	Nutrient removal	998

Table 19. Llama 3.3 70B output for *Chlorella* genus

Genus	Temperature	N _{trial}	Order	Output Concept	Score
<i>Dunaliella</i>	0.0	1	1	Microalgae cultivation	8398
			2	Lipid production	6893
			3	Biomass production	5585
			4	Growth rate optimization	4555
			5	Protein production	3731
			6	Wastewater treatment	3574
			7	Carotenoid production	2826
			8	Fatty acid production	2450
			9	Chlorophyll research	2272
			10	Salinity effects	2243
		1	1	Microalgae cultivation	8398
			2	Lipid production	6893
			3	Biomass production	6553
			4	Aquaculture research	5595
			5	Wastewater treatment	4575
			6	Photosynthesis research	4555
			7	Marine ecology	4144
			8	Bioproduct development	3731
			9	Protein production	3629
			10	Carbon capture	3574
		2	1	Microalgae cultivation	8398
			2	Lipid production	6893
			3	Biomass production	6553
			4	Algal biotechnology	5595
			5	Wastewater treatment	3574
			6	Aquaculture feed production	3355
			7	Photosynthesis research	3178
			8	Marine ecology	2944
			9	Protein production	2826
			10	Carbon capture	2450
	0.1	3	1	Microalgae cultivation	8398
			2	Lipid production	6893
			3	Biomass production	6553
			4	Growth rate optimization	6231
			5	Species identification	5733
			6	Protein production	5595
			7	Wastewater treatment	5273
			8	Carotenoid production	5039
			9	Fatty acid production	4555
			10	Photosynthesis research	4481
		4	1	Microalgae cultivation	8398
			2	Lipid production	6893
			3	Biomass production	6553
			4	Growth rate optimization	6231
			5	Protein production	5595
			6	Wastewater treatment	5273
			7	Carotenoid production	4920
			8	Aquaculture feed production	4555
			9	Photosynthesis research	4429
			10	Marine ecology	4170
		5	1	Microalgae cultivation	8398
			2	Lipid production	6893
			3	Algal biomass production	5595
			4	Microalgae species identification	5288
			5	Wastewater treatment	3574
			6	Aquaculture feed production	2762
			7	Photosynthesis research	2450
			8	Marine ecology	2243
			9	Bioproduct development	2217
			10	Protein production	2115

Table 20. Llama 3.3 70B output for *Dunaliella* genus

Genus	Temperature	N _{trial}	Order	Output Concept	Score
<i>Galdieria</i>	0.0	1	1	Microalgae cultivation	58
			2	Bioproduct production	36
			3	Wastewater treatment	31
			4	Phycocyanin production	25
			5	Cell growth medium	29
		1	1	Microalgae cultivation	58
			2	Bioproduct production	36
			3	Wastewater treatment	31
			4	Phycocyanin production	25
			5	Biomass production	21
			6	Nutrient utilization	22
			7	Cell growth medium	20
			8	Glucose utilization	17
			9	Species cultivation	15
			10	Microalgae biomass	14
	0.1	2	1	Microalgae cultivation	58
			2	Bioproduct production	36
			3	Wastewater treatment	31
			4	Phycocyanin production	25
			5	Biomass production	21
			6	Nutrient utilization	22
			7	Cell growth medium	29
			8	Glucose utilization	21
		3	1	Microalgae cultivation	58
			2	Bioproduct production	36
			3	Wastewater treatment	31
			4	Phycocyanin production	25
			5	Biomass production	21
			6	Nutrient utilization	22
			7	Cell growth medium	29
			8	Glucose utilization	21
		4	1	Microalgae cultivation	58
			2	Bioproduct production	36
			3	Wastewater treatment	31
			4	Phycocyanin production	25
			5	Biomass production	21
			6	Nutrient utilization	22
			7	Cell growth medium	29
			8	Glucose utilization	21
		5	1	Microalgae cultivation	58
			2	Bioproduct production	36
			3	Wastewater treatment	31
			4	Phycocyanin production	25
			5	Biomass production	21
			6	Nutrient utilization	22
			7	Cell growth medium	20
			8	Glucose utilization	21
			9	Species cultivation	36
			10	Microalgae production	64

Table 21. Llama 3.3 70B output for *Galdieria* genus

Genus	Temperature	N _{trial}	Order	Output Concept	Score
<i>Chlorella</i>	0.0	1	1	Microalgae biochemical composition and photosynthesis	42063.0
			2	Microalgae cultivation for lipid/biofuel production	35619.0
			3	Microalgae-based wastewater treatment and nutrient removal	17558.0
			4	Process optimization and modeling of microalgae production	14189.0
			5	Biofuel and bioproduct development	2133.0
		1	1	Microalgae cultivation for biomass and lipid/biofuel production	41123.0
			2	Microalgae biochemical composition and photosynthesis (carotenoids, proteins, etc.)	27081.0
			3	Microalgae-based wastewater treatment and nutrient removal	19305.0
			4	Process optimization and extraction yield improvement	15870.0
			5	Biofuel production from microalgae	2133.0
	0.1	2	1	Biomass production	14734.0
			2	Wastewater treatment	9982.0
			3	Lipid production for biodiesel/biofuel	9090.0
			4	Cultivation technology (photobioreactor, culture media)	8955.0
			5	Energy efficiency and cost performance	5193.0
			6	Photosynthesis and light optimization	3615.0
			7	Protein production	1438.0
		3	1	Microalgae cultivation and growth optimization	20458.0
			2	Wastewater treatment and nutrient removal using microalgae	9157.0
			3	Lipid production for biodiesel and biofuel	8282.0
			4	Biomass production for feed and food applications	5369.0
			5	Pigment (carotenoid/chlorophyll) production and photosynthesis research	1540.0
		4	1	Microalgae cultivation and biomass production	33213.0
			2	Photobioreactor technology and process optimization	9927.0
			3	Wastewater treatment and nutrient removal using microalgae	9147.0
			4	Lipid accumulation for biodiesel and biofuel production	9090.0
			5	Protein production for feed applications	1438.0
		5	1	Microalgae biology, photosynthesis and protein production	43706.0
			2	Microalgae cultivation for biomass and lipid/biodiesel production	28147.0
			3	Wastewater treatment and nutrient removal using microalgae	18513.0
			4	Extraction, processing and optimization technologies for biofuel production	13525.0

Table 22. GPT OSS 120B output for *Chlorella* genus

Genus	Temperature	N _{trial}	Order	Output Concept	Score
<i>Dunaliella</i>	0.0	1	1	Microalgae cultivation for biomass production	37990.0
			2	Cellular response and treatment studies	28615.0
			3	Bioproduct development and applications	19034.0
			4	Lipid and carotenoid production for biofuel	18056.0
			5	Species and strain selection in microalgae	10813.0
		1	1	Pigment and lipid biosynthesis	1102.46
			2	Biomass production and bioproduct cultivation	757.58
			3	Abiotic factors influencing microalgal growth	617.26
			4	Strain-specific microalgae studies	415.48
			5	Experimental treatment approaches	194.19
	0.1	2	1	Microalgae cultivation for biomass and bioproducts	1583.0
			2	Growth optimization under light, nitrogen, and salinity	1084.0
			3	Pigment production (carotenoids and chlorophyll)	515.0
			4	Activity and treatment effects on microalgae	455.0
			5	Lipid and fatty acid production for biofuel	344.0
		3	6	Species-specific studies (Chlorella, Tetraselmis)	199.0
			1	Microalgae growth and nutrient optimization	32883.0
			2	Biomass production and cultivation parameters	28611.0
			3	Pigment and lipid bioactive compound synthesis	12461.0
			4	Species-specific strain studies	9768.0
			5	Bioactivity and treatment effects	4916.0
		4	1	Growth conditions and physiological responses	3766.0
			2	Cultivation and biomass production	3029.0
			3	Lipid and pigment production	1572.0
			4	Species and strain selection	1262.0
			5	Treatment effects on activity	772.0
		5	1	Microalgae cultivation and growth optimization	4631.0
			2	Bioproduct development and biomass production	3749.0
			3	Pigment and lipid biosynthesis in microalgae	1572.0
			4	Species-specific microalgae research	900.0
			5	Bioactivity and treatment effects	897.0

Table 23. GPT OSS 120B output for *Dunaliella* genus

Genus	Temperature	N _{trial}	Order	Output Concept	Score
<i>Galdieria</i>	0.0	1	1	Phycocyanin production via microalgae cultivation	2040.0
			2	Algal biomass production from wastewater	933.0
		1	1	Phycocyanin production from microalgae cultivation	2594.0
			2	Wastewater-based microalgae biomass production for bioproducts	2123.0
	0.1	2	1	Microalgae phycocyanin production	2391.0
			2	Wastewater-based algal biomass production	1133.0
		3	1	Microalgae cultivation & phycocyanin production	2391.0
			2	Wastewater-driven algal biomass production	1133.0
		4	1	Phycocyanin production from microalgae	340.3
			2	Wastewater-driven microalgae biomass production	141.3
		5	1	Phycocyanin production via microalgae cultivation	2391.0
			2	Wastewater-based microalgae biomass production	1133.0

Table 24. GPT OSS 120B output for *Galdieria* genus

2. Screening for the minimal number of occurrences to retain a keyword

When creating the network associated with a genus, it is important to choose the minimal number of occurrences to retain a keyword with care. Indeed, it is very intuitive to feel that a one-size-fits-all value does not exist. To get a grasp of it, one only needs to consider the *Galdieria* genus, with 123 articles, and *Chlorella* genus with 9771 articles. Setting a fixed occurrence, for example 50, would lead to a too stringent selection for the first one, while being too broad for the second. This first comment allows to realize that an adequate occurrence level is surely a function of the number of articles featuring a given genus. Consequently, several scaling methods were tested with several genera. Three types of scaling were opted for for this preliminary screening: linear, square root, and logarithmic, with three levels of modulation. In terms of genera, as discussed above, the number of articles featuring the genus matters. Therefore, five genera, covering the whole spectrum, were selected: *Chlorella* (9771 articles), *Scenedesmus* (3318 articles), *Dunaliella* (1451 articles), *Chaetoceros* (798 articles), and *Galdieria* (128 articles). While this procedure was initially unsuccessful, it allowed for the generation of several networks (Fig. 1 to 15), amounting to enough data to produce a power law fit of the minimum number of occurrences to obtain a satisfying network versus the number of articles retained in the corpus.

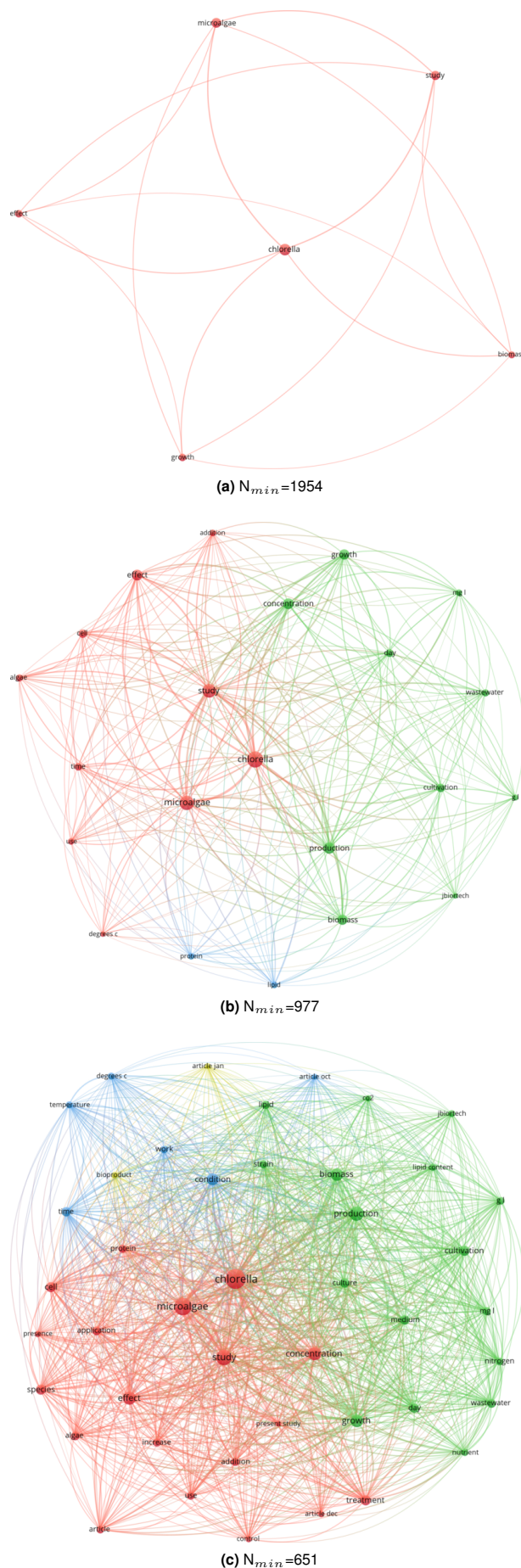
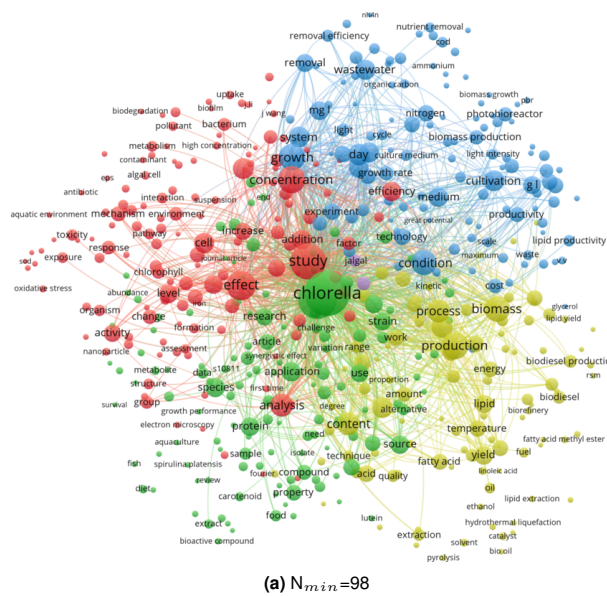
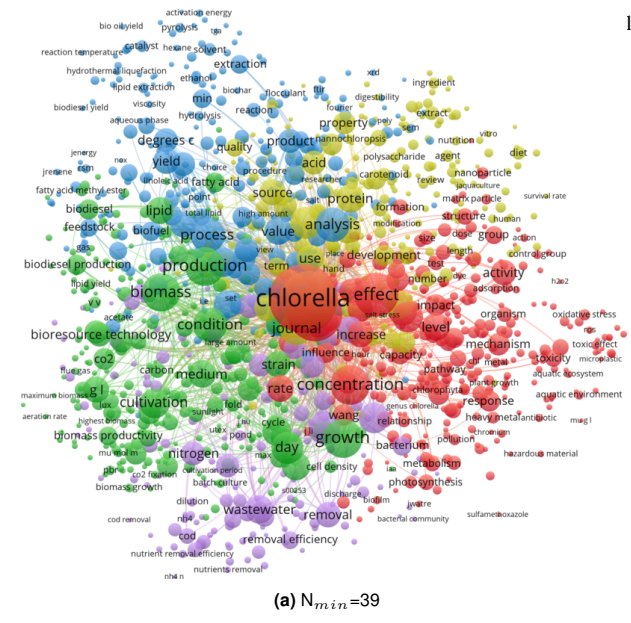


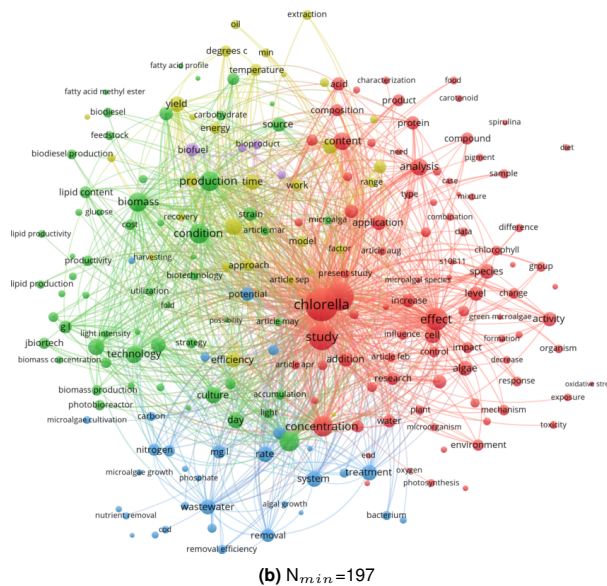
Fig. 1. Networks obtained with VOSviewer for different values of the minimum occurrence of a keyword to be retained. Genus: *Chlorella* (9771 articles), linear scaling



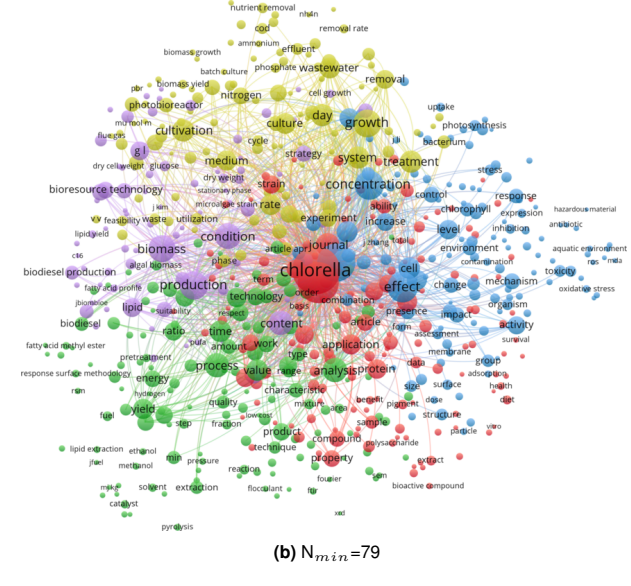
(a) $N_{min}=98$



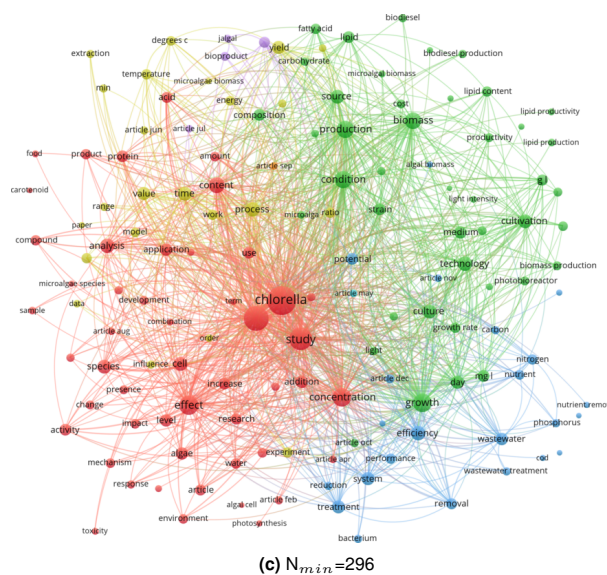
(a) $N_{min}=39$



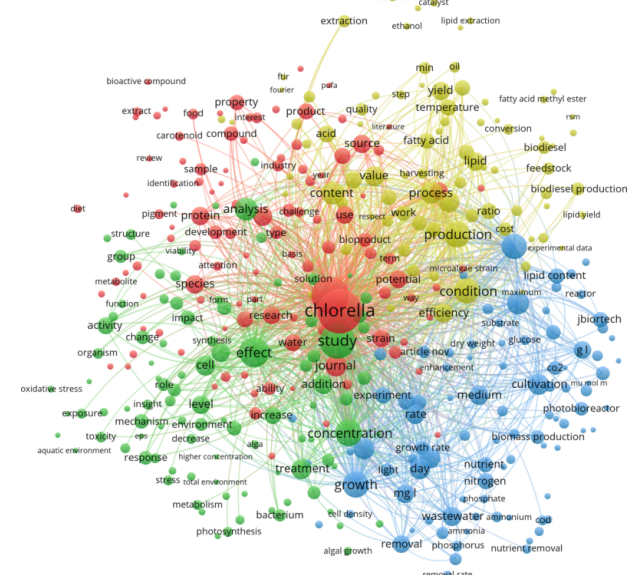
(b) $N_{min}=197$



(b) $N_{min}=79$



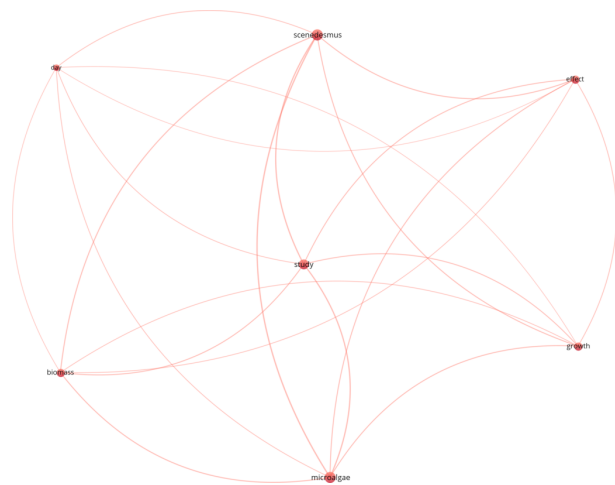
(c) $N_{min}=296$



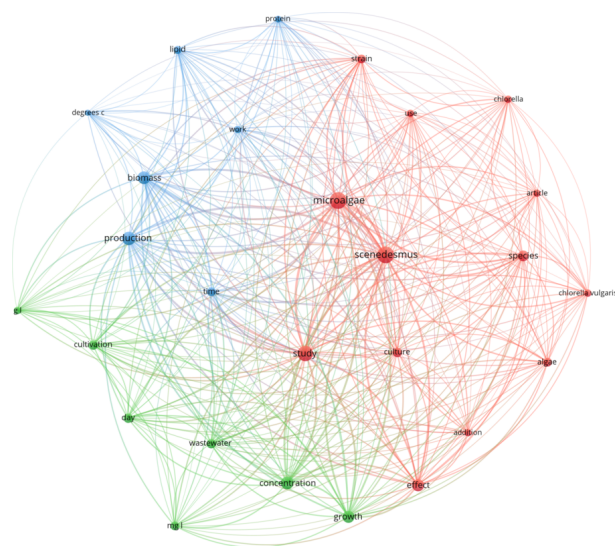
(c) $N_{min}=119$

Fig. 2. Networks obtained with VOSviewer for different values of the minimum occurrence of a keyword to be retained. Genus: *Chlorella* (9771 articles), square root scaling

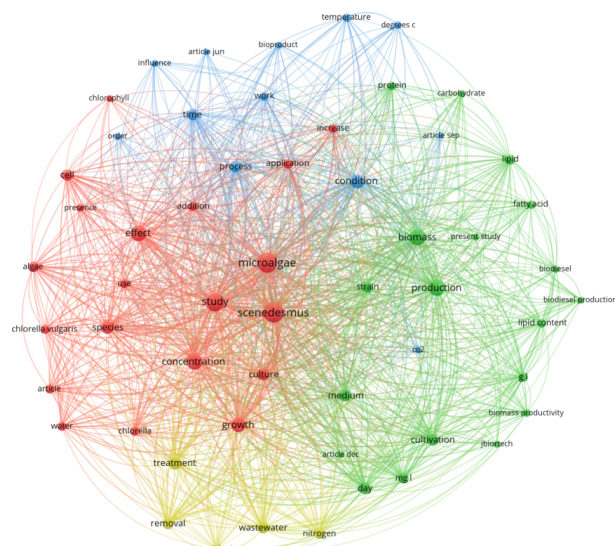
Fig. 3. Networks obtained with VOSviewer for different values of the minimum occurrence of a keyword to be retained. Genus: *Chlorella* (9771 articles), logarithmic (base 10) scaling



(a) $N_{min}=663$

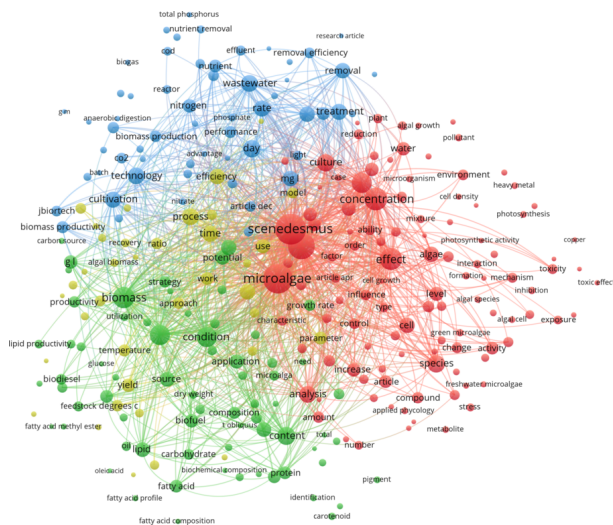


(b) $N_{min}=331$

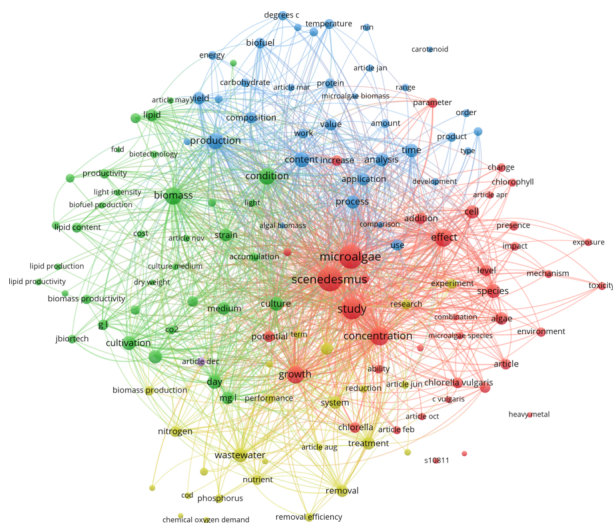


(c) $N_{min}=221$

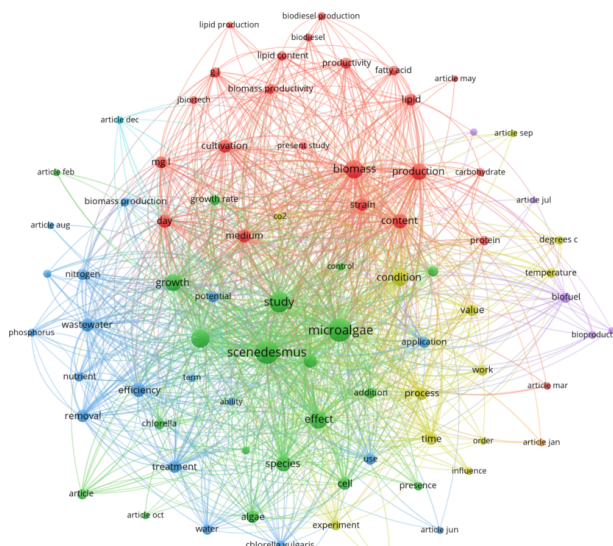
Fig. 4. Networks obtained with VOSviewer for different values of the minimum occurrence of a keyword to be retained. Genus: *Scenedesmus* (3318 articles), linear scaling



(a) $N_{min}=57$



(b) $N_{min}=115$



(c) $N_{min}=172$

Fig. 5. Networks obtained with VOSviewer for different values of the minimum occurrence of a keyword to be retained. Genus: *Scenedesmus* (3318 articles), square root scaling

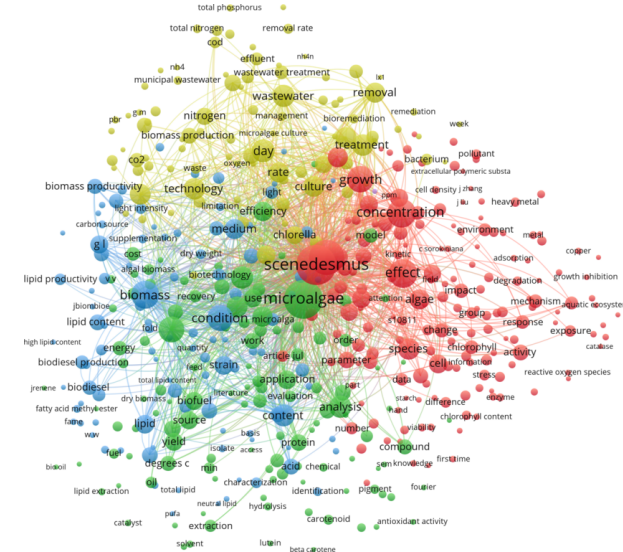
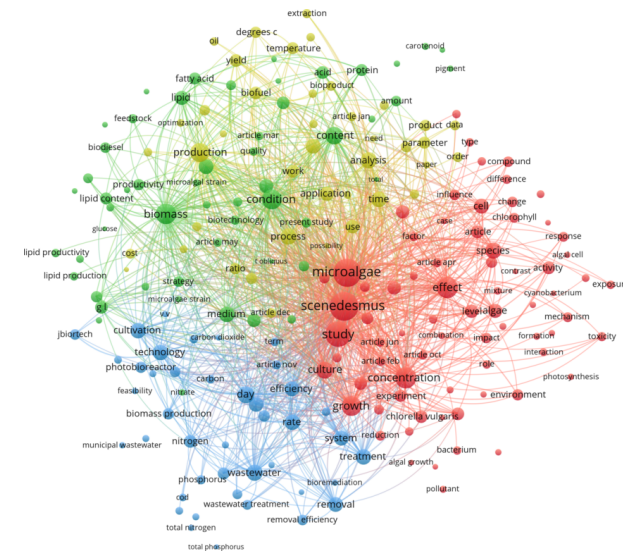
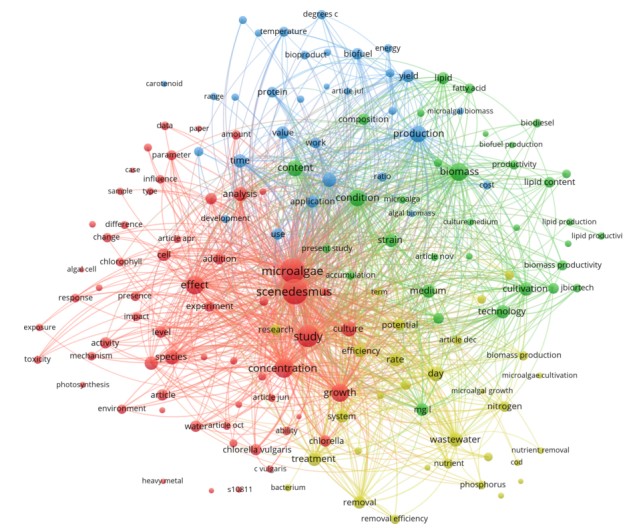
(a) $N_{min}=35$ (b) $N_{min}=70$ (c) $N_{min}=105$

Fig. 6. Networks obtained with VOSviewer for different values of the minimum occurrence of a keyword to be retained. Genus: *Scenedesmus* (3318 articles), logarithmic (base 10) scaling

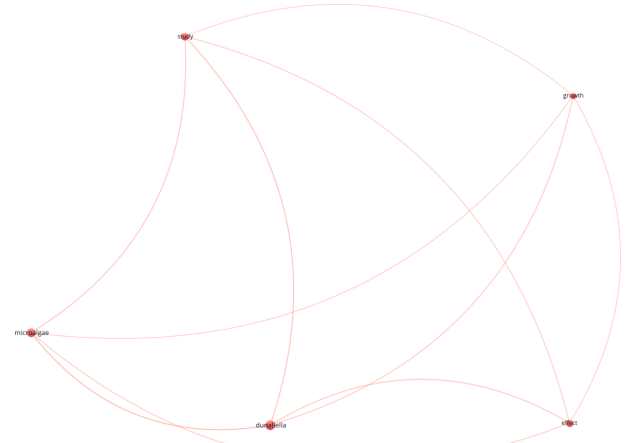
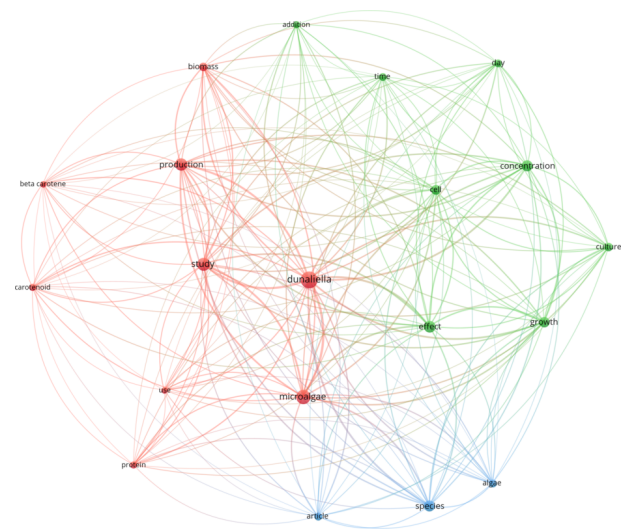
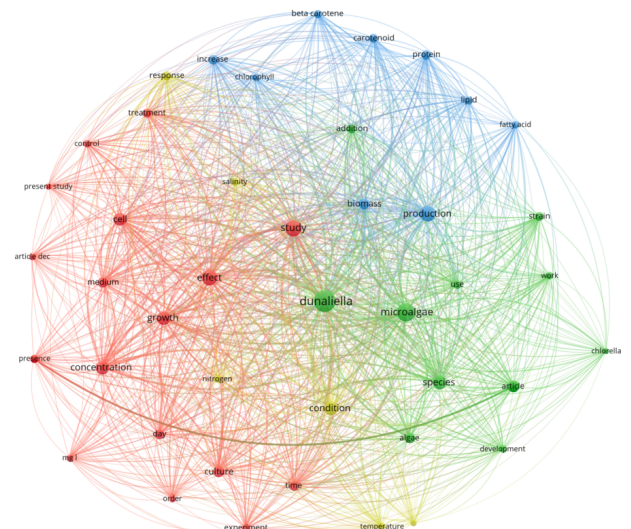
(a) $N_{min}=290$ (b) $N_{min}=145$ (c) $N_{min}=96$

Fig. 7. Networks obtained with VOSviewer for different values of the minimum occurrence of a keyword to be retained. Genus: *Dunaliella* (1451 articles), linear scaling

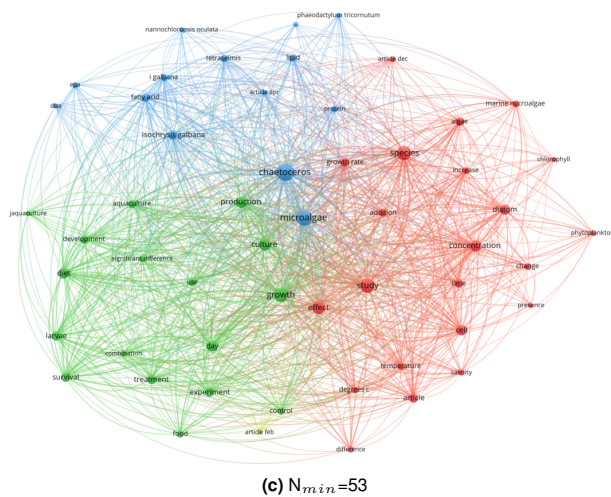
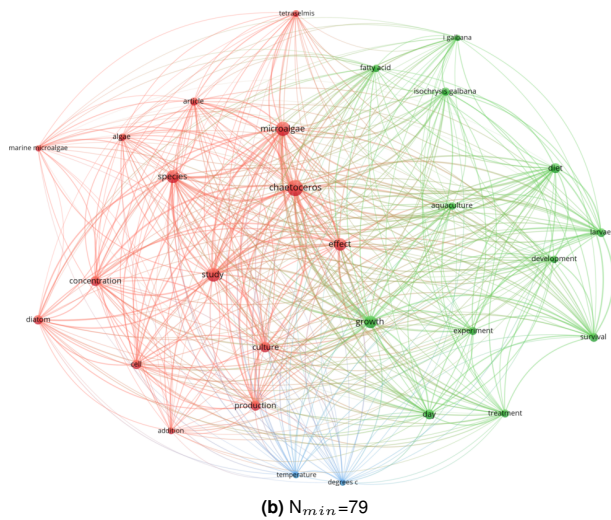


Fig. 10. Networks obtained with VOSviewer for different values of the minimum occurrence of a keyword to be retained. Genus: *Chaetoceros* (798 articles), linear scaling

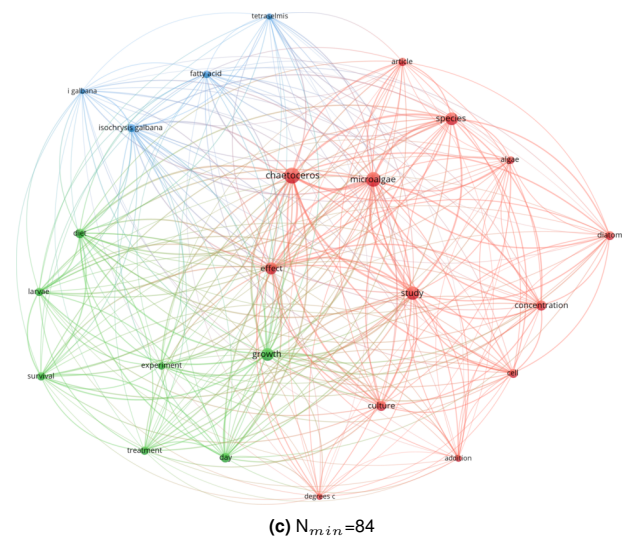
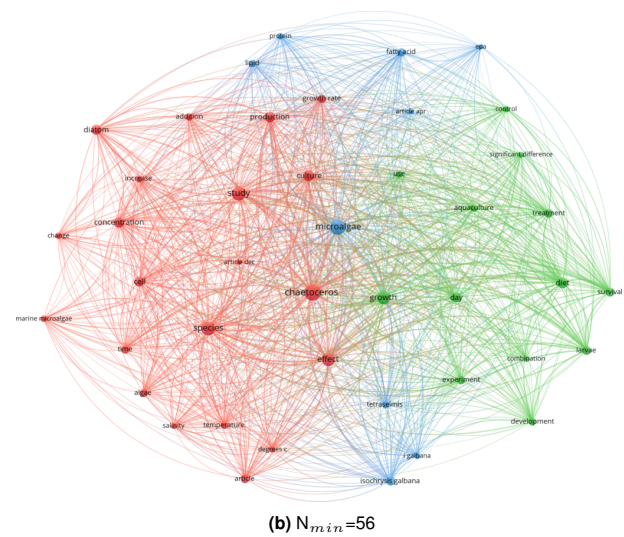
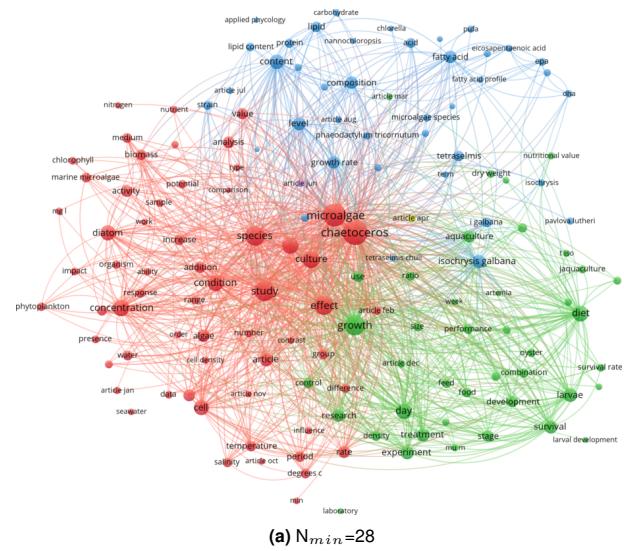
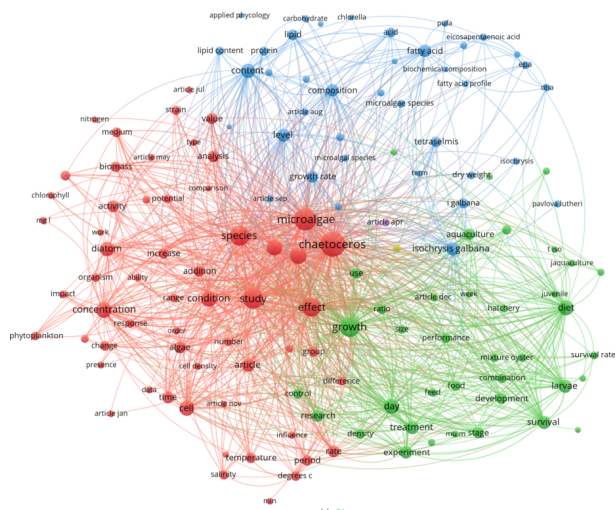
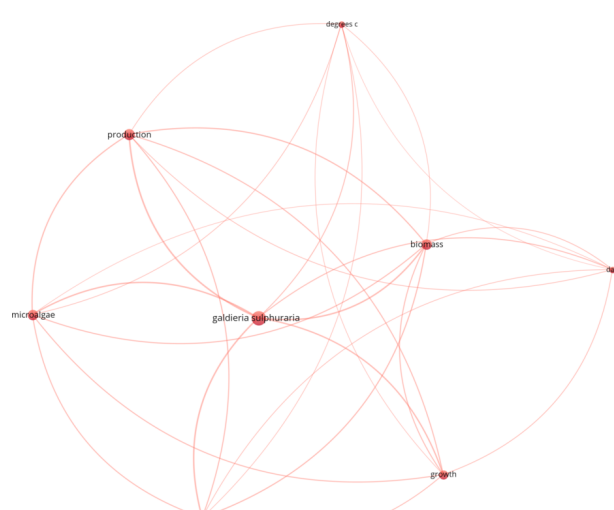


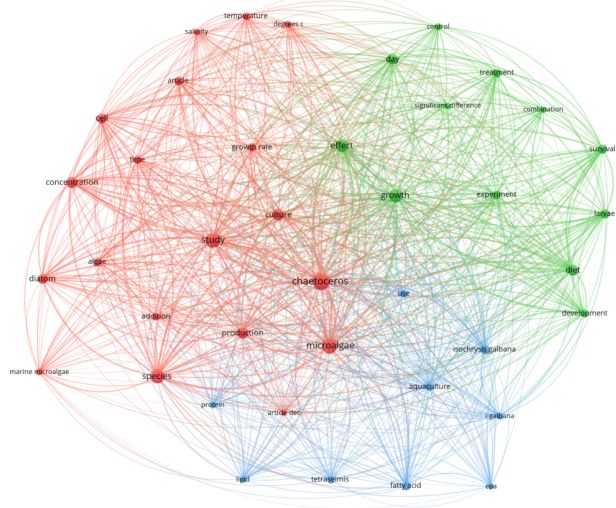
Fig. 11. Networks obtained with VOSviewer for different values of the minimum occurrence of a keyword to be retained. Genus: *Chaetoceros* (798 articles), square root scaling



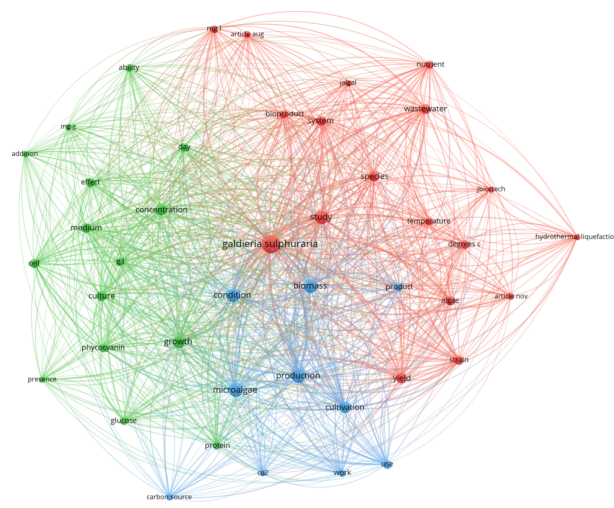
(a) $N_{min}=29$



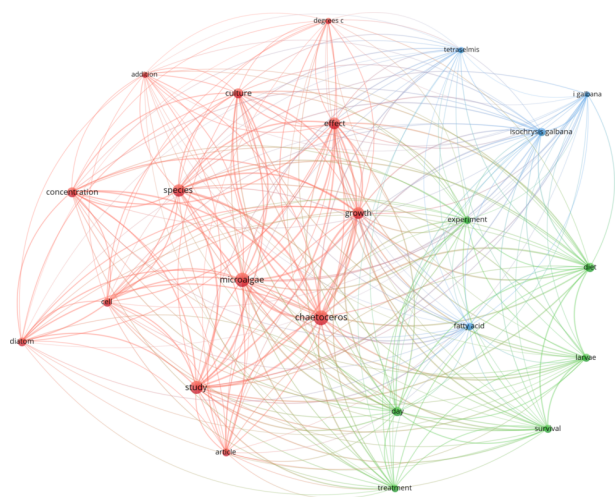
(a) $N_{min}=25$



(b) $N_{min}=58$

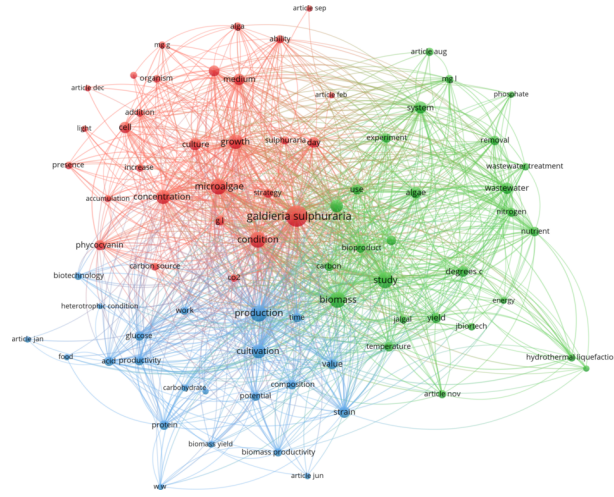


(b) $N_{min}=12$



(c) $N_{min}=87$

Fig. 12. Networks obtained with VOSviewer for different values of the minimum occurrence of a keyword to be retained. Genus: *Chaetoceros* (798 articles), logarithmic (base 10) scaling



(c) $N_{min}=8$

Fig. 13. Networks obtained with VOSviewer for different values of the minimum occurrence of a keyword to be retained. Genus: *Galdieria* (128 articles), linear scaling

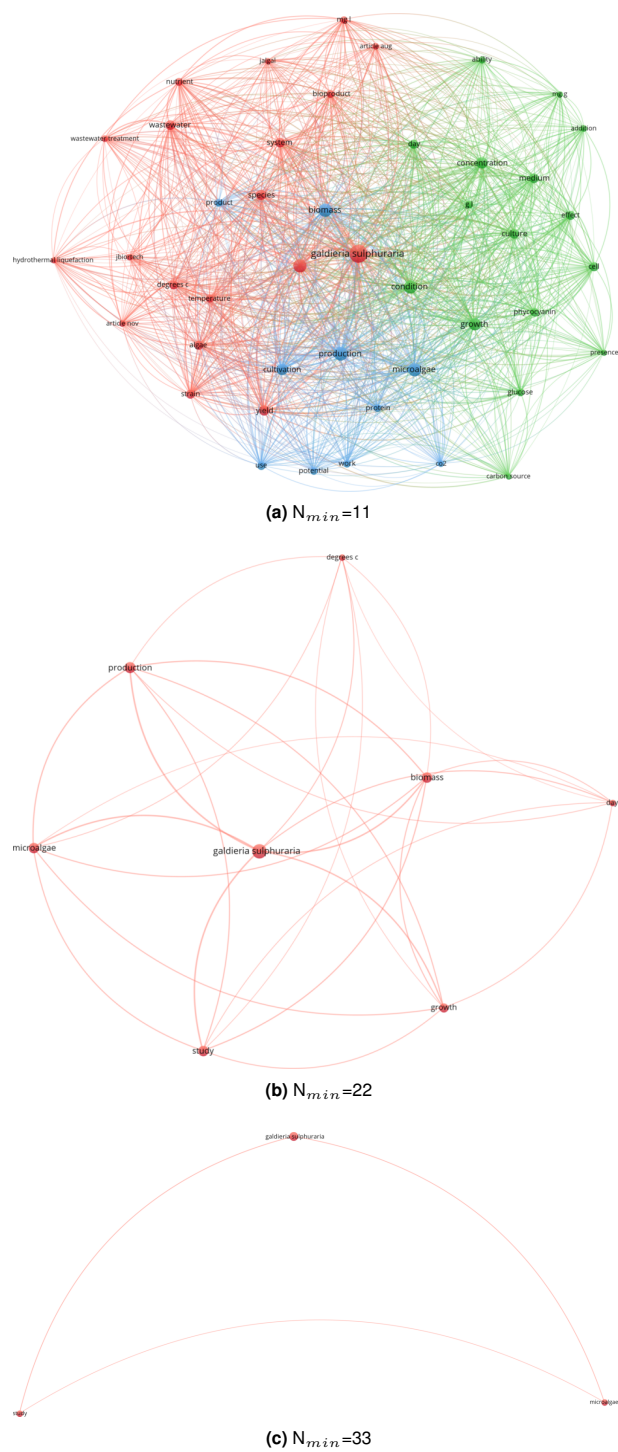


Fig. 14. Networks obtained with VOSviewer for different values of the minimum occurrence of a keyword to be retained. Genus: *Galdieria* (128 articles), square root scaling

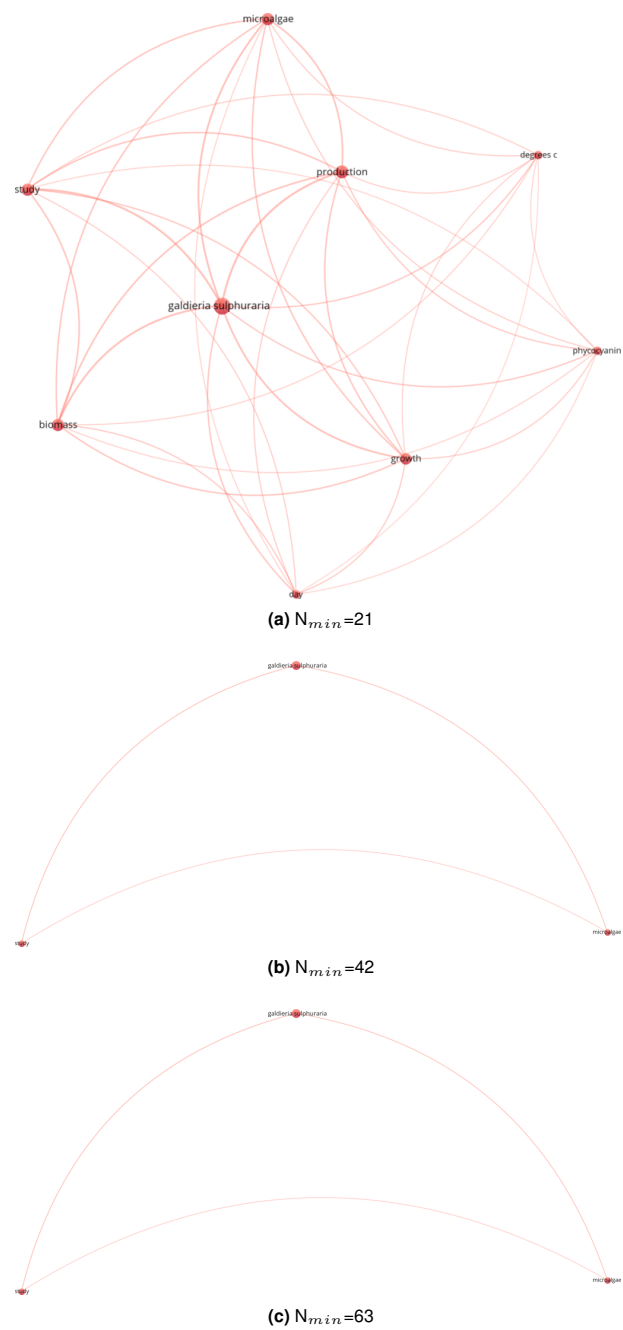


Fig. 15. Networks obtained with VOSviewer for different values of the minimum occurrence of a keyword to be retained. Genus: *Galdieria* (128 articles), logarithmic (base 10) scaling

3. Prompt engineering

Three prompting styles were tested with different levels of intensity in the directives given to the model. In all cases, the prompts were designed following a clear structure: presentation of the task, instructions to carry out the tasks, and examples of the expected outputs (few-shot alike approach). The three prompts are reported below using *Galdieria* genus as an example (owing to the limited length of the network passed to the model).

3.1. Basis prompt

YOUR TASK: UNDERLYING CONCEPT IDENTIFICATION FROM WEIGHTED WORD CLUSTERS

You are a bibliometric network analysis assistant.

You will be given a CSV file representing a co-occurrence network.

Each line will be formatted as: Keyword,Cluster,Occurrences,Strength

Your task will be to extract and rank the MAIN RESEARCH CONCEPTS underlying in the network.

INSTRUCTIONS (FOLLOW EXACTLY)

STEP 1 — REMOVE NOISE KEYWORD

Ignore keywords that are:

- Measurement units (e.g., mg l, g l, degrees c, etc)
- Generic research words (e.g., study, use, day, etc)
- Very broad terms (e.g., microalgae, study, etc)
- Journal names or abbreviations (e.g., jbiotech, jalg, etc)
- Publication dates or fragments (e.g., article nov, article jan, etc)

Do NOT remove biologically meaningful terms.

STEP 2 — INITIAL CLUSTER GROUPING

Use the provided "cluster" field as the initial grouping.

Each cluster is a candidate concept.

You can enrich each of the candidate cluster with keywords from the other clusters.

Enrichment must remain marginal.

STEP 3 — REFINE CONCEPTS

Refine the cluster to identify the MAIN RESEARCH CONCEPTS they represent.

The underlying concepts are often supported by uncommon words. For example:

- Quantum yield
- Pyrolysis
- Removal rate

Use a FEW WORDS to describe each concept. For example:

- Aquaculture feed production
- Larvae feed production
- Lipid production for biofuel
- Lipid production for feed
- Wastewater treatment
- Photosynthesis research

- Marine ecology

- Protein production

- Carbon capture and sequestration

"Occurrences" and "Strength" are unlikely to be useful at this step.

STEP 4 — REFINE CONCEPTS

Rank the identified concept based on perceived importance.

Perceived importance can be approached using Occurrences, Strength, Word count of the constituting the cluster.

OUTPUT FORMAT (STRICT)

1 ||Concept_1||Score_1

2 ||Concept_2||Score_2

No explanations.

No commentary.

No extra text.

NOW ANALYZE THIS CSV FILE:

Keyword,Cluster,Occurrence,Strength

biomass,2,58,393

bioproduct,2,21,150

cell,1,29,164

condition,1,58,358

cultivation,1,44,297

day,2,26,191

degrees c,2,26,151

g l,1,20,153

glucose,1,21,139

growth,1,48,324

medium,1,31,217

mg l,2,17,125

microalgae,1,58,351

nutrient,2,22,159

phycocyanin,1,25,164

production,1,64,417

species,2,36,200

study,2,58,360

use,1,25,163

wastewater,2,31,231

This prompt output can be found in Tables 19, 20, and 21 and will be used as references to compare the next prompts to.

3.2. Permissive prompt

YOUR TASK: UNDERLYING CONCEPT IDENTIFICATION FROM WEIGHTED WORD CLUSTERS

You are a bibliometric network analysis assistant.

You will be given a CSV file representing a co-occurrence network.

Each line will be formatted as: Keyword,Cluster,Occurrences,Strength

Your task will be to extract and rank the main research concepts underlying in the network.

INSTRUCTIONS

STEP 1 — REMOVE NOISE KEYWORD

As a hint to can preferentially remove keywords unlikely to bear meaningful information, such as:

- Measurement units (e.g., mg l, g l, degrees c, etc)
- Generic research words (e.g., study, use, day, etc)
- Very broad terms (e.g., microalgae, study, etc)
- Journal names or abbreviations (e.g., jbiotech, jalgol, etc)
- Publication dates or fragments (e.g., article nov, article jan, etc)

STEP 2 — IDENTIFYING CONCEPTS

Cluster the keywords to identify the main research concepts they represent.

The underlying concepts are often supported by uncommon words. For example:

- Quantum yield
- Pyrolysis
- Removal rate

Use a few words to describe each concept. For example:

- Aquaculture feed production
- Larvae feed production
- Lipid production for biofuel
- Lipid production for feed
- Wastewater treatment
- Photosynthesis research
- Marine ecology
- Protein production
- Carbon capture and sequestration

STEP 3 — RANK CONCEPTS

Rank the identified concept based on perceived importance.

Perceived importance can be approached using Occurrences, Strength, Word count of the constituting the cluster.

OUTPUT FORMAT (STRICT)

1 ||Concept_1||Score_1

2 ||Concept_2||Score_2

NOW ANALYZE THIS CSV FILE:

Keyword,Cluster,Occurrence,Strength

biomass,2,58,393

bioproduct,2,21,150

cell,1,29,164
 condition,1,58,358
 cultivation,1,44,297
 day,2,26,191
 degrees c,2,26,151
 g l,1,20,153
 glucose,1,21,139
 growth,1,48,324
 medium,1,31,217
 mg l,2,17,125
 microalgae,1,58,351
 nutrient,2,22,159
 phycocyanin,1,25,164
 production,1,64,417
 species,2,36,200
 study,2,58,360
 use,1,25,163
 wastewater,2,31,231

The first noticeable difference with this prompt is that the model struggled to adhere to the requested output. Often, it supplied explanations along with the identified concepts. Detailed outputs can be found in Tables 25, 26, and 27. As one can see, the model opted for a formatting detaching from the examples proposed by the prompt. Still, the lower guidance led to the model missing important elements such as β -carotene for *Dunaliella* genus and phycocyanin for *Galdieria* genus. Furthermore, it reduced the proposed concepts for *Galdieria* genus, leading to an oversimplification of the network.

Genus	Temperature	N _{trial}	Order	Output Concept	Score
<i>Chlorella</i>	0.0	1	1	Lipid production	0.85
			2	Biomass production	0.78
			3	Wastewater treatment	0.74
			4	Biofuel production	0.69
			5	Photosynthesis research	0.64
			6	Microalgae cultivation	0.61
			7	Carbon capture	0.57
			8	Protein production	0.54
			9	Biotechnology	0.51
			10	Environmental impact	0.48
	0.1	1	1	Lipid Production	2341.0
			2	Biomass Production	2011.0
			3	Wastewater Treatment	1832.0
			4	Biofuel Production	1742.0
			5	Photosynthesis Research	1456.0
			6	Microalgae Cultivation	1364.0
			7	Carbon Capture	1245.0
			8	Protein Production	1142.0
			9	Biotechnology	1063.0
			10	Environmental Impact	943.0
	0.1	4	1	Lipid Production	2341.0
			2	Biomass Production	2011.0
			3	Wastewater Treatment	1832.0
			4	Biofuel Production	1742.0
			5	Photosynthesis Research	1431.0
			6	Microalgae Cultivation	1356.0
			7	Carbon Capture	1245.0
			8	Protein Production	1142.0
			9	Biotechnology	1063.0
			10	Environmental Impact	943.0
	0.1	3	1	Lipid production	0.85
			2	Biomass production	0.78
			3	Wastewater treatment	0.74
			4	Biofuel production	0.69
			5	Photosynthesis research	0.64
			6	Microalgae cultivation	0.59
			7	Nutrient removal	0.55
			8	Carbon capture	0.51
		4	1	Biomass Production	0.85
			2	Lipid Production	0.78
	0.1	5	3	Wastewater Treatment	0.74
			4	Biofuel Production	0.69
			5	Microalgae Cultivation	0.65
			6	Photosynthesis Research	0.59
			7	Carbon Capture	0.56
			8	Protein Production	0.53
			9	Biotechnology	0.49
			10	Environmental Impact	0.45
			1	Biomass Production	2341.0
			2	Lipid Production	1945.0
			3	Wastewater Treatment	1832.0
			4	Biofuel Production	1734.0
			5	Microalgae Cultivation	1649.0
			6	Photosynthesis Research	1542.0
			7	Carbon Capture	1463.0
			8	Protein Production	1438.0
			9	Biotechnology	1360.0
			10	Environmental Impact	1291.0

Table 25. Llama 3.3 70B Instruct output for *Chlorella* genus with the permissive prompt

Genus	Temperature	N _{trial}	Order	Output Concept	Score
<i>Dunaliella</i>	0.0	1	1	Lipid production	nan
			2	Aquaculture feed production	nan
			3	Photosynthesis research	nan
			4	Wastewater treatment	nan
			5	Marine ecology	nan
			6	Bioproduct development	nan
			7	Growth and cultivation	nan
			8	Nutrient content	nan
		1	1	Lipid production	0.85
			2	Aquaculture and feed production	0.78
			3	Bioproduct development	0.74
			4	Photosynthesis and growth	0.72
			5	Wastewater treatment	0.69
			6	Environmental factors	0.66
			7	Cellular response	0.64
			8	Protein production	0.62
		2	1	Lipid production	0.85
			2	Aquaculture feed production	0.78
			3	Marine ecology	0.74
			4	Bioproduct development	0.69
			5	Cell culture	0.65
			6	Protein production	0.62
			7	Wastewater treatment	0.59
			8	Photosynthesis research	0.55
	0.1	3	1	Lipid production	0.85
			2	Aquaculture feed production	0.78
			3	Wastewater treatment	0.69
			4	Photosynthesis research	0.64
			5	Marine ecology	0.59
			6	Bioproduct development	0.55
			7	Cell culture	0.52
			8	Protein production	0.49
		4	1	Lipid production	0.85
			2	Aquaculture feed production	0.78
			3	Biomass production	0.72
			4	Marine ecology	0.65
			5	Wastewater treatment	0.59
			6	Bioproduct development	0.55
			7	Protein production	0.52
			8	Photosynthesis research	0.48
		5	1	Lipid production for biofuel	0.85
			2	Aquaculture feed production	0.78
			3	Wastewater treatment	0.73
			4	Photosynthesis research	0.69
			5	Marine ecology	0.65
			6	Bioproduct development	0.62
			7	Growth and cultivation	0.59
			8	Nutrient and environmental effects	0.55

Table 26. Llama 3.3 70B Instruct output for *Dunaliella* genus with the permissive prompt

Genus	Temperature	N _{trial}	Order	Output Concept	Score
<i>Galdieria</i>	0.0	1	1	Microalgae cultivation	1241.0
			2	Bioproduct production	821.0
			3	Wastewater treatment	567.0
	0.1	1	1	Microalgae cultivation	1241.0
			2	Bioproduct production	814.0
			3	Wastewater treatment	390.0
			4	Cell growth and production	744.0
		2	1	Microalgae cultivation	1241.0
			2	Bioproduct production	821.0
			3	Wastewater treatment	390.0
			4	Cell growth and production	734.0
		3	1	Microalgae cultivation	0.53
			2	Bioproduct production	0.29
			3	Wastewater treatment	0.18
		4	1	Microalgae cultivation	1241.0
			2	Bioproduct production	814.0
			3	Wastewater treatment	390.0
			4	Cell growth and production	744.0
		5	1	Microalgae cultivation	1241.0
			2	Bioproduct production	821.0
			3	Wastewater treatment	431.0

Table 27. Llama 3.3 70B Instruct output *Galdieria* genus with the permissive prompt

3.3. Directive prompt

YOUR TASK: UNDERLYING CONCEPT IDENTIFICATION FROM WEIGHTED WORD CLUSTERS

You are a bibliometric network analysis expert.

You will be given a CSV file representing a co-occurrence network.

Each line will be formatted as: Keyword,Cluster,Occurrences,Strength

Your task will be to extract and rank the main research concepts underlying in the network.

Do not hesitate to think amply to improve the identification of the concepts.

Pay attention to details as some important concepts may be obfuscated by other, therefore, extensive reasoning can be needed.

INSTRUCTIONS (FOLLOW EXACTLY)

STEP 1 — REMOVE NOISE KEYWORD

Ignore keywords that are:

- Measurement units (e.g., mg l, g l, degrees c, etc)
 - Generic research words (e.g., study, use, day, etc)
 - Very broad terms (e.g., microalgae, study, etc)
 - Journal names or abbreviations (e.g., jbiotech, jalgol, etc)
 - Publication dates or fragments (e.g., article nov, article jan, etc)
 - ANY OTHER KEYWORDS that would find "too common"
- Do NOT remove biologically meaningful terms.*

STEP 2 — INITIAL CLUSTER GROUPING

Use the provided "cluster" field as the INITIAL grouping (you can latter detach from the initial seeds).

Each cluster is a candidate concept.

You can enrich each of the candidate cluster with keywords from the other clusters.

Enrichment must remain marginal.

STEP 3 — REFINE CONCEPTS

Refine the cluster to identify the MAIN RESEARCH CONCEPTS they represent.

You can now ignore the initial grouping.

The underlying concepts are often supported by uncommon words. FOR EXAMPLE:

- Quantum yield
- Pyrolysis
- Removal rate

(only use the above examples if needed, you can use your own wording)

Use a FEW WORDS to describe each concept. For example:

- Aquaculture feed production
- Larvae feed production
- Lipid production for biofuel
- Lipid production for feed
- Wastewater treatment
- Photosynthesis research

- Marine ecology

- Protein production

- Carbon capture and sequestration

(only use the above examples if needed, you can use your own wording)

"Occurrences" and "Strength" are unlikely to be useful at this step.

STEP 4 — RANK CONCEPTS

Rank the identified concept based on perceived importance.

Perceived importance can be approached using Occurrences, Strength, Word count of the constituting the cluster.

OUTPUT FORMAT (STRICT)

1 ||Concept_1||Score_1

2 ||Concept_2||Score_2

No explanations.

No commentary.

No extra text.

NOW ANALYZE THIS CSV FILE:

Keyword,Cluster,Occurrence,Strength

biomass,2,58,393

bioproduct,2,21,150

cell,1,29,164

condition,1,58,358

cultivation,1,44,297

day,2,26,191

degrees c,2,26,151

g l,1,20,153

glucose,1,21,139

growth,1,48,324

medium,1,31,217

mg l,2,17,125

microalgae,1,58,351

nutrient,2,22,159

phycocyanin,1,25,164

production,1,64,417

species,2,36,200

study,2,58,360

use,1,25,163

wastewater,2,31,231

With this prompting style, the model adheres to the requested output formatting. Still, the variations from the regular prompt are minimal and adversarial (Tables 28, 29, and 30). Indeed, despite being pushed towards deeper thinking, the model loses some of the subtlety it was able to bring to the concept identification (as detailed in the body of the manuscript).

Genus	Temperature	N _{trial}	Order	Output Concept	Score
Chlorella	0.0	1	1	Microalgae cultivation	6244
			2	Biomass production	2724
			3	Biofuel production	1325
			4	Wastewater treatment	1653
			5	Lipid production	1421
			6	Photosynthesis research	449
			7	Biotechnology applications	706
			8	Microalgal biomass production	544
			9	Nutrient removal	998
			10	Carbon capture	684
		1	1	Microalgae cultivation	6244
			2	Biomass production	2724
			3	Biofuel production	1325
			4	Wastewater treatment	1653
			5	Lipid production	1421
			6	Photosynthesis research	449
			7	Biotechnology applications	706
			8	Algal biomass production	432
			9	Nutrient removal	998
			10	Carbon capture	684
		2	1	Microalgae cultivation	6244
			2	Biomass production	2724
			3	Biofuel production	1325
			4	Wastewater treatment	1653
			5	Lipid production	1421
			6	Photosynthesis research	449
			7	Biotechnology applications	706
			8	Microalgal biomass production	544
			9	Nutrient removal	998
			10	Carbon capture	684
	0.1	3	1	Microalgae cultivation	6244
			2	Biomass production	2724
			3	Biofuel production	1325
			4	Wastewater treatment	1653
			5	Lipid production	1421
			6	Photosynthesis research	449
			7	Microalgae species	406
			8	Biotechnology	706
			9	Carbon capture	684
			10	Nutrient removal	404
		4	1	Microalgae cultivation	6244
			2	Biomass production	2724
			3	Biofuel production	1325
			4	Wastewater treatment	1653
			5	Lipid production	1421
			6	Photosynthesis research	449
			7	Microalgae species	406
			8	Biotechnology	706
			9	Carbon capture	684
			10	Nutrient removal	404
		5	1	Microalgae cultivation	6244
			2	Biomass production	2724
			3	Biofuel production	1325
			4	Wastewater treatment	1653
			5	Lipid production	1421
			6	Photosynthesis research	449
			7	Biotechnology applications	706
			8	Algal biomass production	432
			9	Nutrient removal	998
			10	Carbon capture	684

Table 28. Llama 3.3 70B Instruct output for *Chlorella* genus with the directive prompt

Genus	Temperature	N _{trial}	Order	Output Concept	Score
<i>Dunaliella</i>	0.0	1	1	Microalgae cultivation	8398
			2	Lipid production	6893
			3	Biomass production	6553
			4	Growth rate optimization	5585
			5	Carotenoid production	4920
			6	Wastewater treatment	3574
			7	Protein production	3178
			8	Marine microalgae cultivation	2762
			9	Bioproduct application	2450
			10	Photosynthesis research	2243
		1	1	Microalgae cultivation	8398
			2	Lipid production	6893
			3	Biomass production	5585
			4	Algal growth rate	5288
			5	Carotenoid production	4555
			6	Wastewater treatment	3731
			7	Protein production	3537
			8	Marine microalgae	3317
			9	Bioproduct application	3178
			10	Photosynthesis research	2974
		2	1	Microalgae cultivation	8398
			2	Lipid production	6893
			3	Biomass production	5585
			4	Algal species research	5288
			5	Photosynthesis and growth	4555
			6	Bioproduct application	3731
			7	Wastewater treatment	3574
			8	Carotenoid production	2450
			9	Protein production	2528
			10	Marine ecology	2318
	0.1	3	1	Microalgae cultivation	8398
			2	Lipid production	6893
			3	Biomass production	3731
			4	Algal species diversity	5288
			5	Photosynthesis and growth	5585
			6	Bioproduct application	2336
			7	Wastewater treatment	2192
			8	Carotenoid production	2450
			9	Fatty acid synthesis	1767
			10	Protein production	2528
		4	1	Microalgae cultivation	8398
			2	Lipid production	6893
			3	Biomass production	5585
			4	Algal growth rate	5288
			5	Carotenoid production	4555
			6	Wastewater treatment	3731
			7	Protein production	3536
			8	Marine microalgae	3317
			9	Bioproduct application	3178
			10	Photosynthesis research	2974
		5	1	Microalgae cultivation	8398
			2	Lipid production	6893
			3	Biomass production	5585
			4	Algal growth rate	5288
			5	Carotenoid production	4555
			6	Wastewater treatment	3731
			7	Protein production	3537
			8	Marine microalgae	3317
			9	Bioproduct application	3178
			10	Photosynthesis research	2974

Table 29. Llama 3.3 70B Instruct output for *Dunaliella* genus with the directive prompt

Genus	Temperature	N _{trial}	Order	Output Concept	Score
<i>Galdieria</i>	0.0	1	1	Microalgae cultivation	417
			2	Bioproduct production	393
			3	Wastewater treatment	358
			4	Microalgae growth	324
			5	Phycocyanin production	217
			6	Nutrient management	200
			7	Biomass production	164
			8	Cell growth medium	153
		1	1	Microalgae cultivation	58
			2	Bioproduct production	36
			3	Wastewater treatment	31
			4	Phycocyanin production	25
			5	Biomass production	22
			6	Nutrient utilization	21
			7	Cell growth medium	20
		2	1	Microalgae cultivation	58
			2	Bioproduct production	36
			3	Wastewater treatment	31
			4	Phycocyanin production	25
			5	Biomass production	22
			6	Nutrient utilization	21
			7	Cell growth medium	20
	0.1	3	1	Microalgae cultivation	58
			2	Bioproduct production	36
			3	Wastewater treatment	31
			4	Phycocyanin production	25
			5	Biomass production	22
			6	Nutrient utilization	21
			7	Cell growth medium	20
		4	1	Microalgae cultivation	58
			2	Bioproduct production	36
			3	Wastewater treatment	31
			4	Phycocyanin production	25
			5	Biomass production	22
			6	Nutrient utilization	21
			7	Cell growth medium	20
			8	Glucose utilization	17
		5	1	Microalgae cultivation	58
			2	Bioproduct production	36
			3	Wastewater treatment	31
			4	Phycocyanin production	25
			5	Biomass production	22
			6	Nutrient utilization	21
			7	Cell growth medium	20

Table 30. Llama 3.3 70B Instruct output *Galdieria* genus with the directive prompt

4. Concepts identification results & Domain associations

Once the workflow has been calibrated, it was run for the 48 genera identified using VOSviewer. The results are contained within Tables 31 and 78.

Concept	Domain(s)
Microalgae cultivation	1
Lipid production	3
Biomass production	1
Carbohydrate increase	1
Growth effect	1
Species strain	2
Temperature effect	1
Aquaculture feed	6

Table 31. Concepts identified for *Acutodesmus* genus

Concept	Domain(s)
Microalgae growth and production	1 & 11
Dinoflagellate species and microalgae	2 & 10
Fatty acid production and concentration	3 & 4
Marine microalgae and growth effect	1 & 11
Cell and species presence	2 & 10

Table 32. Concepts identified for *Amphidinium* genus

Concept	Domain(s)
Microalgae biomass production	1
Diatom culture for lipid production	3
Microalgae species growth	1
Navicula and Nitzschia species research	2 & 11
Protein production from microalgae	4
Amphora coffeaeformis and microalgae interaction	10 & 11

Table 33. Concepts identified for *Amphora* genus

Concept	Domain(s)
Microalgae Growth	1
Biomass Production	1
Cyanobacterium Culture	1
Algal Species Research	2
Nitrogen Effect	1
Chlorella Vulgaris Strain	2
Microalgae Concentration	1
Cell Growth Optimization	1
Water Quality Research	8
Lipid Production	3

Table 34. Concepts identified for *Anabaena* genus

Concept	Domain(s)
Microalgae production	1
Algal biomass production	1
Lipid production	3
Microalgae growth	1
Species and strain selection	2
Culture medium optimization	1
Chlorella research	11
Wastewater treatment	8
Algal lipid content	3

Table 35. Concepts identified for *Ankistrodesmus* genus

Concept	Domain(s)
Microalgae cultivation	1
Protein production	4
Lipid production	3
Wastewater treatment	8
Bioactive compound extraction	4
Carotenoid production	7
Biomass production	1
CO2 utilization	9
Aquaculture feed production	6
Phycoremediation	8

Table 36. Concepts identified for *Arthrospira* genus

Concept	Domain(s)
Lipid production	3
Fatty acid production	3
DHA production	4
Biomass production	1
Microalgae cultivation	1
Aquaculture feed	6
Growth optimization	1
Strain development	2
Wastewater treatment	8

Table 37. Concepts identified for *Aurantiochytrium* genus

Concept	Domain(s)
Microalgae cultivation	1
Biofuel production	3
Biomass production	1
Lipid production	3
Hydrocarbon production	3
Protein production	4
Chlorella vulgaris cultivation	1
Biodiesel production	3
Algal growth	1
Culture medium optimization	1

Table 38. Concepts identified for *Botryococcus* genus

Concept	Domain(s)
Microalgae culture	1
Aquaculture feed production	6
Lipid production	3
Marine microalgae research	11
Wastewater treatment	8
Protein production	4
Microalgae growth rate	1
Fatty acid production	3
Larvae feed production	6
Algal biochemistry	11

Table 39. Concepts identified for *Chaetoceros* genus

Concept	Domain(s)
Microalgae production	1
Photosynthesis research	11
Lipid production	3
Biotechnology application	1
Wastewater treatment	8
Biofuel production	3
Protein production	4
Aquaculture feed production	6
Carbon capture	9
Algal bioproducts	4

Table 40. Concepts identified for *Chlamydomonas* genus

Concept	Domain(s)
Microalgae cultivation	1
Biomass production	1
Biofuel production	3
Wastewater treatment	8
Lipid production	3
Growth rate	1
Photosynthesis research	11
Biotechnology	1
Carbon capture	9
Nutrient removal	8

Table 41. Concepts identified for *Chlorella* genus

Concept	Domain(s)
Microalgae production	1
Biomass production	1
Lipid production	3
Wastewater treatment	8
Algal growth	1
Fatty acid production	3
Biodiesel production	3
Microalgae strain	2
Microalgae species	2
Chlorella production	1

Table 42. Concepts identified for *Chlorococcum* genus

Concept	Domain(s)
Microalgae research	11
Lipid production	3
Biomass growth	1
Fatty acid production	3
Algal species study	2
Microalgal strain development	2
Algal condition optimization	1
Cell growth effect	11

Table 43. Concepts identified for *Coccomyxa* genus

Concept	Domain(s)
Microalgae Growth	1
Algal Biomass Production	1
Lipid Production	3
Strain Development	2
Culture Medium Optimization	1
Protein Production	4
Species Identification	2
Bioproduct Development	4
Chlorella Research	11
Coelastrella Sp Research	2

Table 44. Concepts identified for *Coelastrella* genus

Concept	Domain(s)
Microalgae growth	1
Diatom species	2
Cell culture	1
Chlorophyll concentration	7
Biomass production	1
Cylindrotheca fusiformis	2
Photosynthesis research	11
Aquatic species study	2

Table 45. Concepts identified for *Cylindrotheca* genus

Concept	Domain(s)
Microalgae cultivation	1
Biodiesel production	3
Biomass production	1
Lipid production	3
Wastewater treatment	8
Bioproduct development	4
Algal strain improvement	2
Nutrient effects on growth	1
CO2 utilization	9
Fatty acid production	3

Table 46. Concepts identified for *Desmodesmus* genus

Concept	Domain(s)
Microalgae cultivation	1
Lipid production	3
Algal biomass production	1
Microalgae growth rate	1
Wastewater treatment	8
Carotenoid production	7
Bioproduct application	4
Aquaculture feed production	6
Protein production	4
Photosynthesis research	11

Table 47. Concepts identified for *Dunaliella* genus

Concept	Domain(s)
Microalgae biomass production	1
Cell growth and concentration	1
Algal species and presence	2 & 10
Protein and paramylon production	4 & 5
Microalgae growth effect	1

Table 48. Concepts identified for *Euglena* genus

Concept	Domain(s)
Microalgae cultivation	1
Bioproduct production	4
Wastewater treatment	8
Phycocyanin production	7
Biomass production	1
Nutrient management	1
Cell growth medium	1
Glucose utilization	11

Table 49. Concepts identified for *Galdieria* genus

Concept	Domain(s)
Microalgae cultivation	1
Astaxanthin production	7
Lipid production	3
Biomass production	1
Carotenoid production	7
Fatty acid production	3
Protein production	4
Wastewater treatment	8
Bioproduct development	4
Microalgae growth	1

Table 50. Concepts identified for *Haematococcus* genus

Concept	Domain(s)
Algal Bloom Research	10
Microalgae Culture	1
Algal Growth Factors	1
Harmful Algal Bloom	10
Microalgae Species	2

Table 51. Concepts identified for *Heterosigma* genus

Concept	Domain(s)
Microalgae production	1
Aquaculture feed production	6
Fatty acid composition	1
Lipid production	3
Growth rate optimization	1
Biomass production	1
Protein production	4
Wastewater treatment	8
Marine microalgae cultivation	1
Chaetoceros calcitran cultivation	1

Table 52. Concepts identified for *Isochrysis* genus

Concept	Domain(s)
Cyanobacteria growth	1
Lipid production	3
Microalgae strain	2
Biomass concentration	1

Table 53. Concepts identified for *Leptolyngbya* genus

Concept	Domain(s)
Microalgae production	1
Microalgae growth	1
Lipid production	3
Biomass production	1
Chlorella strain	2
Nitrogen effect	1
Microalgae lipid content	3

Table 54. Concepts identified for *Micractinium* genus

Concept	Domain(s)
Microalgae production	1
Algal bloom	10
Water treatment	8
Cyanobacteria growth	1
Biomass production	1
Chlorophyll content	7
Microcystin removal	8
Nitrogen presence	1
Species strain	2
Temperature effect	1

Table 55. Concepts identified for *Microcystis* genus

Concept	Domain(s)
Microalgae cultivation	1
Lipid production	3
Biomass production	1
Biodiesel production	3
Bioproduct development	4
Strain development	2
Fatty acid production	3
Lipid accumulation	3
Culture optimization	1
Growth control	1

Table 56. Concepts identified for *Monoraphidium* genus

Concept	Domain(s)
Microalgae cultivation	1
Biofuel production	3
Lipid production	3
Biomass production	1
Wastewater treatment	8
Protein production	4
Biodiesel production	3
Carbon capture	9
Algal oil extraction	3
Microalgae species identification	2

Table 57. Concepts identified for *Nannochloropsis* genus

Concept	Domain(s)
Microalgae growth	1
Diatom species	2
Algal biomass	1
Culture conditions	1
Salinity effects	1
Temperature influence	1
Benthic diatom abundance	10
Microalgae concentration	1
Cell growth rate	1
Species composition	2

Table 58. Concepts identified for *Navicula* genus

Concept	Domain(s)
Microalgae biomass production	1
Lipid production for biofuel	3
Cultivation and growth conditions	1 & 11
Biodiesel production	3
Bioproduct development	4
Algal strain improvement	2
Protein production	4
Nitrogen utilization	11
Carbohydrate production	4
Wastewater treatment	8

Table 59. Concepts identified for *Neochloris* genus

Concept	Domain(s)
Diatom growth	1
Microalgae production	1
Phytoplankton species	2
Algal bloom	10
Biomass production	1
Salinity effect	11
Growth rate increase	1
Chlorophyll concentration	7
Navicula abundance	10
Nitzschia closterium presence	2

Table 60. Concepts identified for *Nitzschia* genus

Concept	Domain(s)
Microalgae strain development	2
Biomass production	1
Cyanobacteria growth	1
Algal culture optimization	1
Nitrogen effect on growth	1
Chlorophyll production	7
Microalgae species identification	2
Wastewater treatment potential	8
Photosynthesis research	11
Lipid production for biofuel	3

Table 61. Concepts identified for *Nostoc* genus

Concept	Domain(s)
Cyanobacterium growth	1
Microalgae biomass	1
Algal species	2
Cyanobacteria strain	2
Biomass concentration	1
Algal growth medium	1
Water presence effect	1
Cell growth study	1

Table 62. Concepts identified for *Oscillatoria* genus

Concept	Domain(s)
Microalgae production	1
Microalgae cultivation	1
Lipid production	3
Biomass growth	1
Bioproduct content	1
Cell culture	1
Nitrogen concentration	1
Protein production	4
Strain selection	2
Wastewater treatment	8

Table 63. Concepts identified for *Parachlorella* genus

Concept	Domain(s)
Aquaculture feed production	6
Microalgae production	1
Fatty acid production	3
Larvae feed production	6
Growth rate optimization	1
Wastewater treatment	8
Photosynthesis research	11
Marine ecology	10
Protein production	4
Carbon capture and sequestration	9

Table 64. Concepts identified for *Pavlova* genus

Concept	Domain(s)
Microalgae cultivation	1
Lipid production	3
Photosynthesis research	11
Bioproduct development	4
Marine ecology	10
Fatty acid production	3
Biomass production	1
Protein production	4
Wastewater treatment	8
Aquaculture feed production	6

Table 65. Concepts identified for *Phaeodactylum* genus

Concept	Domain(s)
Ecotoxicity and Toxicity	10
Microalgae and Aquatic Ecosystem	9 & 10
Growth and Growth Rate	1 & 11
Environmental Impact	10
Algal Species and Concentration	2 & 1
Aquatic Environment and Water	8 & 10
Organism and Ecological Effects	2 & 10
Exposure and Toxic Effects	10
Chlorella Vulgaris and Chlorophyll	2 & 7
Aquaculture and Algal Applications	6 & 4

Table 69. Concepts identified for *Raphidocelis* genus

Concept	Domain(s)
Microalgae biomass production	1
Algal species growth	1
Cyanobacteria strain development	2
Wastewater treatment using microalgae	8
Chlorophyll concentration effects	11
Microalgae cultivation temperature	1
Chlorella and phormidium sp presence	2 & 10
Microalgae-based lipid production	3
Oscillatoria and cyanobacterium research	2 & 11
Algal biomass for biofuel	3

Table 66. Concepts identified for *Phormidium* genus

Concept	Domain(s)
Lipid production	3
Microalgae cultivation	1
Wastewater treatment	8
Biofuel production	3
Biomass production	1
Nutrient removal	8
Microalgal biomass	1
Protein production	4
Chlorella cultivation	1
CO2 utilization	9

Table 70. Concepts identified for *Scenedesmus* genus

Concept	Domain(s)
Microalgae biomass production	1
Lipid production	3
Bioproduct development	4
Microalgae growth and temperature	1 & 11
Strain development	2
Biomass productivity	1
Biotechnology applications	1
Species identification	2
Aquatic ecology	10
Photosynthetic organisms	11

Table 67. Concepts identified for *Picochlorum* genus

Concept	Domain(s)
Lipid production	3
Microalgae production	1
Fatty acid production	3
DHA production	4
Biomass production	1
Aquaculture feed production	6
Wastewater treatment	8
Protein production	4
PUFAs production	4
Schizochytrium limacinum cultivation	1

Table 71. Concepts identified for *Schizochytrium* genus

Concept	Domain(s)
Microalgae biomass production	1
Algal culture and growth	1 & 11
Phycoerythrin and polysaccharide production	7 & 4
Fatty acid and protein production	4 & 5
Microalgae species and concentration	2 & 1
Cell growth and effect of conditions	1
Production and presence of biomass	1
Culture condition and microalgae growth	1
Algal protein and fatty acid production	4 & 5
Microalgae culture and production	1 & 3

Table 68. Concepts identified for *Porphyridium* genus

Concept	Domain(s)
Microalgae culture	1
Algal growth and effect	1 & 11
Phytoplankton and species	2 & 11
Microalgae concentration	1
Diatom and strain	2 & 7
Marine microalgae and toxicity	10 & 11
Algal response and presence	10 & 11
Phaeodactylum tricornutum	2
Diet and addition	4 & 5
Order and change	2

Table 72. Concepts identified for *Skeletonema* genus

Concept	Domain(s)
Microalgae cultivation	1
Cyanobacterium production	1
Microalgae growth	1
Biomass concentration	1
CO2 effects on microalgae	11
Cyanobacteria strain development	2
Microalgae species identification	2
Cultivation medium optimization	1
Chlorophyll production	7
<i>Synechococcus elongatus</i> research	11

Table 73. Concepts identified for *Synechococcus* genus

Concept	Domain(s)
Microalgae photosynthesis	11
Cyanobacterium growth	1
Bioproduct production	4
CO2 culture	9
Gene expression	11
Biomass production	1
Strain development	2
Photosynthesis research	11

Table 74. Concepts identified for *Synechocystis* genus

Concept	Domain(s)
Microalgae cultivation	1
Biomass production	1
Wastewater treatment	8
Lipid production	3
Bioproduct development	4
Microalgae growth	1
Nutrient removal	8
Protein production	4
Biomass composition	1
Culture medium optimization	1

Table 75. Concepts identified for *Tetradasmus* genus

Concept	Domain(s)
Aquaculture and microalgae cultivation	1 & 6
Lipid production for biofuel and feed	3 & 5
Microalgae species and strain development	2
Wastewater treatment and nutrient removal	8
Marine ecology and aquaculture feed production	10 & 6
Photosynthesis research and microalgae growth	11 & 1
Protein production and microalgae biomass	1 & 4
Carbon capture and sequestration through microalgae	9
Larvae feed production and aquaculture development	5 & 6
Microalgae-based food and feed production	4 & 5

Table 76. Concepts identified for *Tetraselmis* genus

Concept	Domain(s)
Microalgae culture and production	1 & 3
Phytoplankton and marine ecology	10 & 2
Algal growth and response to environment	1 & 11
Lipid and fatty acid production	3 & 4
Aquatic species and biodiversity	2 & 10
Photosynthesis and temperature effects	11 & 1
Biotechnology applications and experiments	1
Wastewater treatment and microalgae	8 & 1
Cell growth and concentration	1
Protein production and microalgae	4 & 5

Table 77. Concepts identified for *Thalassiosira* genus

Concept	Domain(s)
Microalgae growth and production	1 & 11
Lipid and fatty acid production	3 & 4
Biomass and protein production	1 & 4
Algal culture and strain development	1 & 2
Photosynthesis and temperature effects	11 & 1
Microalgae species and isolate characterization	2
Biotechnological applications and uses	3 & 4
Nutritional content and diet effects	4 & 5

Table 78. Concepts identified for *Tisochrysis* genus

4 1. Bootstrap assessment - Overall workflow robustness

Chlorella (9 771 articles), *Dunaliella* (1 451 articles), and *Galdieria* (128 articles) were selected for workflow robustness analysis using bootstrapping (80% without replacement). Figure 16 presents the average and the confidence interval for the research domains relationships (excluding *Cultivation* for scaling reasons). As one can see, the genus-domain relationships stabilize after approximately 20-30 bootstrap iterations, underlining the robustness of the workflow.

4 2. Refined results - Linear scaling

In the manuscript body, the resulting maps of the genus/domain coupling is displayed using a square root to enhance readability. The same maps, scaled linearly, are available below for the sake of completeness (Fig. 17).

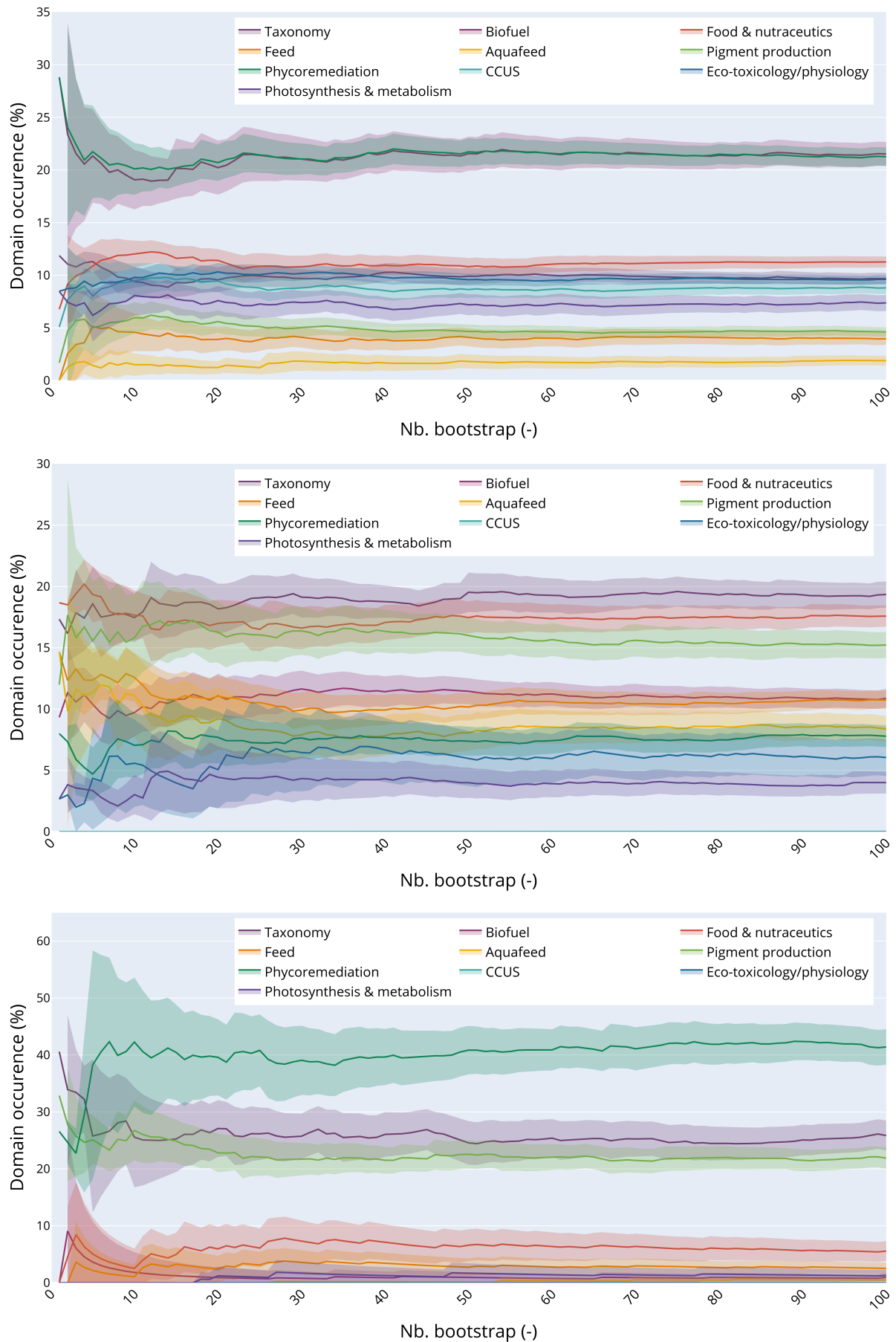


Fig. 16. Average and the confidence interval for the research domains relationships (excluding *Cultivation* for scaling reasons). Top - *Chlorella* genus. Middle - *Dunaliella* genus. Bottom - *Galdieria* genus

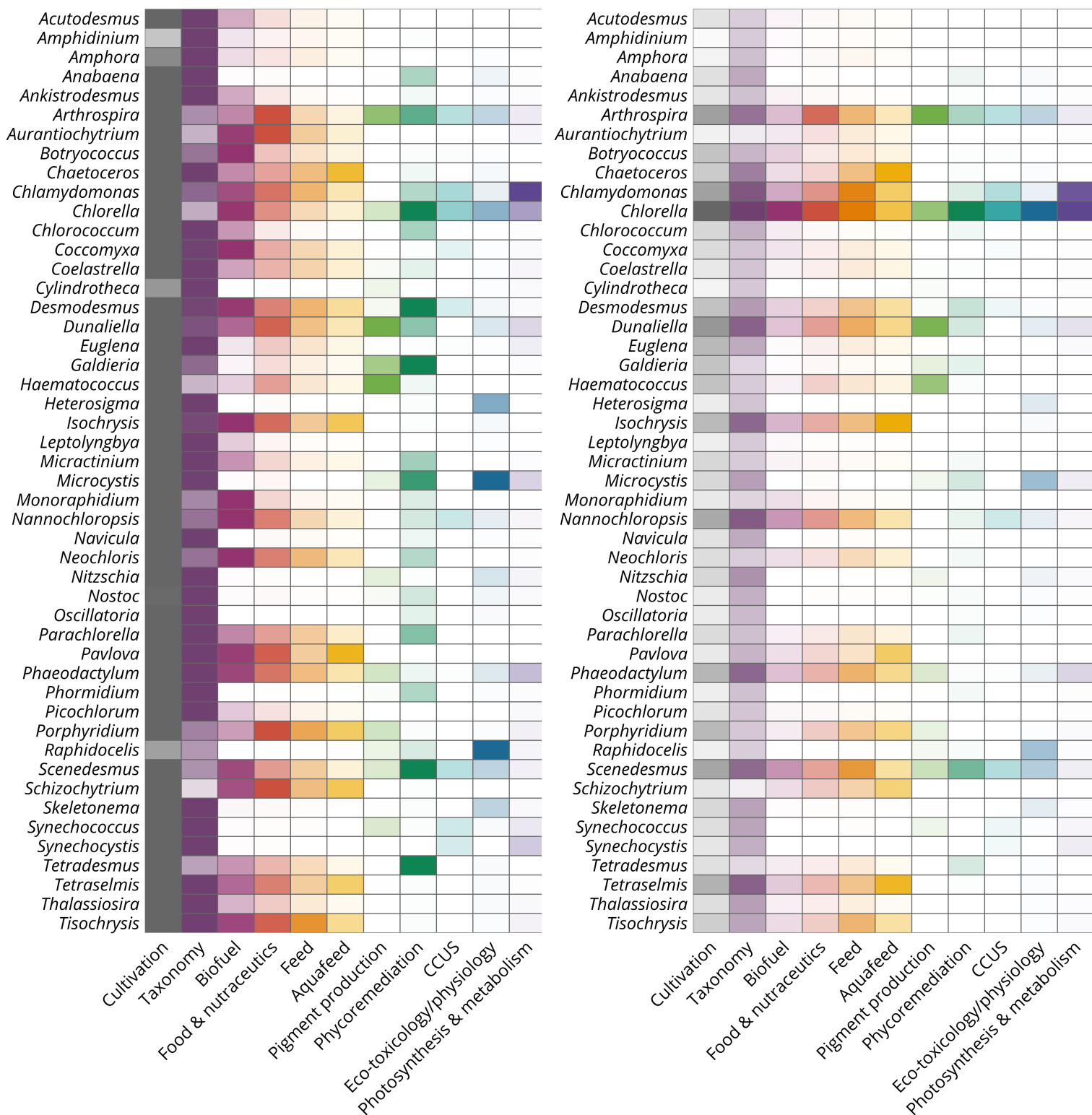


Fig. 17. Association between a specific domain and a genus. Left - genus-wise scaling. Right - domain-wise scaling. Opacity scaled linearly