

Perspective

The CCAP KnowledgeBase: linking protistan and cyanobacterial biological resources with taxonomic and molecular data

CLAIRE M.M. GACHON¹, SVENJA HEESCH^{1,2}, FRITHJOF C. KÜPPER^{1,3}, UNDINE E.M. ACHILLES-DAY¹, DEBRA BRENNAN¹, CHRISTINE N. CAMPBELL¹, ALISON CLARKE¹, RICHARD G. DORRELL⁴, JOANNE FIELD¹, STEVEN GONTAREK¹, CECILIA RAD MENENDEZ¹, RACHEL J. SAXON¹, ANDREA VESZELOVSZKI¹, MICHAEL D. GUIRY², KARIM GHARBI⁵, MARK BLAXTER⁵ & JOHN G. DAY¹

¹The Culture Collection for Algae and Protozoa, Scottish Association for Marine Science, Scottish Marine Institute, Oban PA37 1QA, UK

²AlgaeBase and Irish Seaweed Research Group, Ryan Institute, National University of Ireland Galway, University Road, Galway, Ireland

³Oceanlab, University of Aberdeen, Main Street, Newburgh AB41 6AA, UK

⁴Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA, UK

⁵Edinburgh Genomics, Institute of Evolutionary Biology, Ashworth Laboratories, The King's Buildings, The University of Edinburgh, Edinburgh EH9 3JT, UK

(Received 23 May 2013; revised 31 August 2013; accepted 23 October 2013)

The roles of Biological Resource Centres (BRCs), such as the Culture Collection of Algae and Protozoa (CCAP), have extended beyond their traditional maintenance and provision of curated microorganisms to the user community. A major driver for change has been the exponential increase in metagenomics and environmental sequencing data over the last few years. This has underlined a critical requirement for molecular information on reference biological materials, which would allow better taxonomic interpretation, greater biological understanding and additional exploitation of these data. This is especially relevant for the relatively poorly studied protists (algal/autotrophic as well as heterotrophic) and prokaryotic cyanobacteria, which despite their huge biodiversity, reflected in the genomic data that has been generated, are under-represented in BRCs worldwide. Here we describe the functionalities of the Culture Collection of Algae and Protozoa (CCAP) KnowledgeBase (<http://www.ccap.ac.uk>), developed as a one-stop shop for quality-controlled biological material, hyperlinked to manually curated molecular, bibliographic and taxonomic information. This has been built around the CCAP live collection, which constitutes one of the most genotypically diverse collections in the world with representatives of all the major eukaryotic lineages and the cyanobacteria.

Key words: alga, barcoding, biodiversity, bioinformatics, Biological Resource Centre, cyanobacteria, database, genomics, protist, protozoa

Introduction

The term 'protist' encompasses a phenotypically and genotypically diverse range of mostly unicellular eukaryotic organisms. Many such protistan taxa are capable of photoautotrophic growth, and are generally included as 'algae'; however, many can grow mixotrophically and there are numerous heterotrophic protistan lineages that include free-living, commensal and parasitic organisms. Over the last two decades or so, molecular approaches have helped elucidate their phylogenetic relationships at higher taxonomic levels and revealed that protists include

a much greater diversity of lineages than was previously appreciated (Keeling *et al.*, 2005; Baldauf, 2008; Archibald, 2011). From the perspective of systematics it is reasonable to assert that protists are the most diverse eukaryotic group of organisms, but compared with macroscopic plants and animals, many protistan groups remain poorly characterized. This is not only because of their microscopic size and the lack of obvious defining characters in some taxa, but is also due to difficulties in isolation, cultivation and long-term maintenance. Their diversity has been further substantiated by molecular surveys and metagenomics initiatives that have consistently highlighted the extent of cryptic or unknown protistan

Correspondence to: John Day. E-mail: john.day@sams.ac.uk

diversity (e.g. Massana *et al.*, 2006; Kim *et al.*, 2011). Furthermore, DNA barcoding, whole genome and polyphasic approaches, using both phenotypic and genotypic characters, have shown that many traditional ‘morpho-taxa’ are comprised of a number of phylogenetically distinct clades (Müller *et al.*, 2005; Degerlund *et al.*, 2012; Rybalka *et al.*, 2013).

In addition to their obvious evolutionary biodiversity and phylogenetic importance, autotrophic and heterotrophic protists play key roles in ecosystem function. Algae, including cyanobacteria drive a number of global biogeochemical processes, contributing > 50% of the world’s photosynthetic activity and form the base of the food chain for > 70% of worldwide biomass (Falkowski *et al.*, 2004; Chavez *et al.*, 2011). Similarly, the multifaceted contributions of heterotrophic protists to nutrient cycles is increasingly being recognized including their key roles as consumers (Christaki *et al.*, 1999; Hartmann *et al.*, 2012), degraders/saprotrophs (Kimura *et al.*, 2001) and pathogens (Chambouvet *et al.*, 2008; Gachon *et al.*, 2010). These trophic interactions are clearly crucial to ensuring ecosystem function and stability (Berninger, 1991; Steiner *et al.*, 2006). At the base of pelagic marine food webs, microzooplankton dominate trophic interactions and biogeochemical processes. Thus, their responses to a changing ocean environment have potentially enormous implications for ocean ecosystem functioning (Caron *et al.*, 2013). The term ‘algae’, formerly used in relation to organisms that were *not* something (fungi, liverworts, mosses, ferns, flowering plants, etc.) is now known to encompass a polyphyletic group of organisms that includes some of the smallest (e.g. the prokaryotic cyanophyte *Prochlorococcus*, about 0.5 µm in diameter) and the largest (e.g. the phaeophyceae seaweed *Macrocystis*, reaching 60 m in length) photosynthetic organisms on Earth. Their diverse evolutionary/systematic background is reflected in varied metabolic pathways, e.g. with species of coccolithophores, diatoms and unicellular green algae acquiring iron through various mechanisms (Morrissey & Bowler, 2012), biochemical adaptations such as unique and varied cell-wall carbohydrates (Popper *et al.*, 2011), and an extraordinary halogen metabolism (Küpper *et al.*, 2011). In the future, linking the genetic, molecular and environmental signatures of algae/protists to real specimens that can be manipulated experimentally, or used in mesocosm experimentation, will be crucial to the understanding of the processes involved. Furthermore, the elucidation of functional relationships between different taxa and within species will have implications, not only for blue-skies science/fundamental research, but also for potential applications in manipulating global biogeochemistry (‘geoengineering’) and, less controversially, biotechnology. It is this latter application that is the focus of considerable recent academic and commercial interest (Williams & Laurens, 2010; Milledge, 2011).

Biological Resource Centres (BRCs) such as the Culture Collection for Algae and Protozoa (CCAP) are *de facto* depositories of ‘biological standards’ holding taxonomic and other live reference strains on which much of the associated published science and industrial standards are built (Day & Stacey, 2008). They provide multiple services, ranging from the provision of quality-controlled material for research, commercial and teaching purposes and conservation of patented strains, to the *ex situ* conservation of biodiversity (Gachon *et al.*, 2007). CCAP is the largest European BRC for algae and other protists. The collection comprises more than 2800 publicly available strains: 950 marine algae, 1550 freshwater algae and 300 protozoa. Phylogenetically, the collection is extremely diverse, holding representative taxa from all of the major protistan lineages (Table 1). Over 300 strains are derived from designated taxonomic type material, i.e. the actual strain/specimen to which the scientific name of that organism is formally attached when it was first described in the scientific literature. These cultures may have been subsequently maintained by serial transfer for many decades and may be described as ‘authentic strains’ by some workers (Day *et al.*, 2010). The genotypic and phenotypic stability of these strains in particular are of crucial importance to the scientific community and recommendation on the application of long-term conservation of new holotype specimens, as cryopreserved non-growing specimens, have been made that will ensure their availability to the scientific community (Day *et al.*, 2010).

The provision of cultures from BRCs is of little value unless these cultures are accompanied by information on their identity, provenance and characteristics. Information resources grow with time as work on cultures is published and individual collections, and microorganism database organizations such as StrainInfo at <http://www.straininfo.net/>, will provide more comprehensive data. Here, we focus on the description of the functionalities of the CCAP online repository (<http://www.ccap.ac.uk>), which we have transformed from a simple online catalogue into a one-stop-shop (CCAP KnowledgeBase) for quality-controlled biological material hyperlinked to manually curated molecular, bibliographical and taxonomic information.

System overview

The CCAP online database is used to service the publicly accessible CCAP website and contains a subset of information, principally strain data, synchronized from the restricted CCAP Production database (Fig. 1) when strain data are added or amended. The online database resides on a MySQL database system and also contains the meta-data for the strain-related digital images and videos which are stored on a separate internal file server. Images and

Table 1 Overview of the phylogenetic diversity of CCAP strain holdings¹.

Higher taxonomic level designation ²			Designation used in CCAP KnowledgeBase	No. genera	No. species	No. strains	
Supergroups							
Eubacteria	Cyanobacteria		Cyanophyta	42	93	173	
Eukaryota	Archaeplastida	Rhodophyta	Rhodophyta	46	90	129	
		Glaucophyta	Glaucophyta	1	1	1	
		Chlorophyta	Chlorophyta	189	570	1364	
		Stramenopila	Heterokontophyta (incl. Xanthophyceae, Phaeophyceae, Eustigmatales, Peronosporomycetes & Diatomea)	105	225	678	
	“SAR” (Stramenopiles, Alveolata, Rhizaria)		<i>Incertae sedis</i> (<i>Developayella</i>)	1	1	1	
			Bicosoecida	3	4	6	
		Alveolata	Ciliates/Ciliophora	20	42	61	
			Apicomplexa	1	1	2	
			Dinophyta	13	21	64	
			Cercozoa	Cercozoa	21	36	38
		Excavata	Discoba	Euglenophyta/Euglenozoa	15	56	70
				Heterolobosea	2	2	3
				Jakobea	1	1	1
				Kinetoplastida	2	2	3
	Amoebozoa	Discosea	Discosea	34	88	111	
		Tubulinea	Tubulinea	12	18	23	
	Opisthokonta	Metazoa	Rotifera	1	1	1	
	<i>Incertae sedis</i>	Apusomonadida	Apusomonadida	2	5	5	
		Ancyromonadida	Ancyromonadida	2	7	9	
		Rigifilida	Rigifilida	1	1	1	
		Mantamonadida ³	Mantamonadida	1	1	1	
		Haptophyta	Haptophyta	14	20	30	
		Cryptophyta	Cryptophyta	6	15	30	

¹CCAP Knowledgebase www.ccap.ac.uk interrogated on 12th November 2013. ²Taxonomic designations used by Adl *et al.* (2012) with the addition of Mantamonas/Mantamonadida. ³Glücksman *et al.* (2011).

videos of strains are submitted through a set of internally accessible web pages which are also used to manage the files and maintain the metadata. The forward-facing CCAP website then uses the information in the online database to display strain details and any related images or videos on the relevant webpages, specifically the search result page and the individual strain information pages. Images are watermarked and labelled ‘on the fly’ for web display. Both the internal and external web pages used to display and manage the strain information are written in a mixture of PHP, Javascript and HTML.

Knowledgebase content, organization and features

The KnowledgeBase has grown from the collection’s traditional hard-copy catalogues (e.g. Tompkins *et al.*, 1995) and the later online catalogue, and therefore has a strain-

centric structure. A stable, sequential accession number designates each strain available from the collection. This is associated with a scientific name, which may be updated or amended, for example as the nomenclature evolves, or new studies allow better resolution of the taxonomy (e.g. Pröschold *et al.*, 2001; Smirnov *et al.*, 2011). Information on strain deposition, culturing methods and strain maintenance is easily accessible via dedicated tabs. The KnowledgeBase is updated at least once a month, with key changes or major strain acquisitions highlighted in the ‘News’ section. Users can contact the CCAP by phone, letter, e-mail, or via an online enquiry form and requests are responded to by the CCAP curatorial staff.

Search functions

The KnowledgeBase is fully searchable, using either a simple or an advanced search function. A simple search allows retrieval of strains belonging to a specific taxon, or

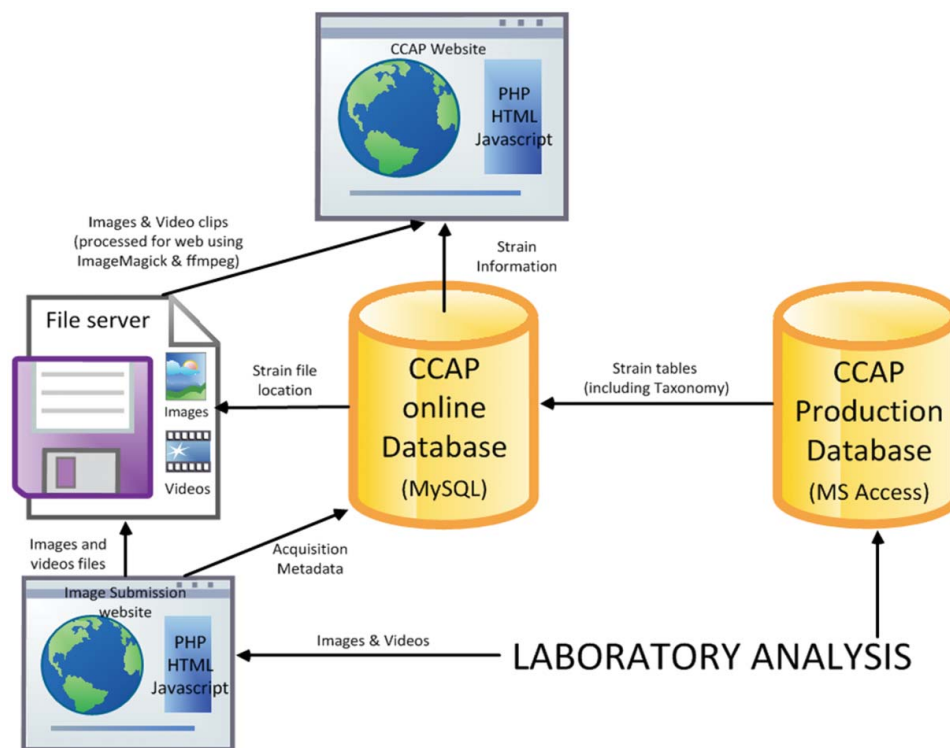


Fig. 1. Overview of the CCAP KnowledgeBase structure

information on strain numbers when the user knows them. The advanced search functionality is especially useful for the identification of related strains belonging to a given taxonomic group, or to design more complex queries based on biogeographical or habitat information. In all queries, the search output is a list of matching entries that the user can further explore individually.

Core strain information

Each strain available from the collection is associated with a standard page record (Fig. 2). It contains core information about the person who isolated the strain, the geographical origin and habitat of the organism, maintenance information, storage method within the CCAP (cryopreservation, serial sub-culture, or both), and possible co-occurring organisms (such as bacteria) in the culture. One field is dedicated to other important information, such as older names and alternative denominations, or whether the strain might be available from other BRCs. Extensive explanations about the scope of each section and the meaning of abbreviations are available in the 'key to strain data' page.

More recently, microscopy pictures, and, when appropriate, videos, have been added to illustrate the typical morphological features of some strains (Fig. 3). Finally, each strain page record is hyperlinked to a list of

references that have been generated using it and hypertext links can take you to individual relevant publications (Fig. 5). This task is on-going, but we are aware of 80–100 peer-reviewed publications generated with CCAP material annually.

Links with other databases

Molecular information: hyperlinking with the European Nucleotide Archive

Until recently, very few molecular marker sequences were available for the organisms held in CCAP; also, the available molecular information was scattered in the literature and was not systematically linked with the source biological material (see Table 1 in Gachon *et al.*, 2007). We have therefore undertaken a systematic programme of generating molecular signatures, from which DNA barcodes can be deduced, of our strain holdings, mostly focusing on obtaining sequences for the small ribosomal subunit and internal transcribed spacers of the ribosomal RNA operon (*rbcL*), with these data made publicly available on the National Center for Biotechnology Information (NCBI)/GenBank, European Nucleotide Archive (ENA) and DNA Data Bank of Japan (DDBJ). Where available, we are also adding sequences derived from CCAP strains that have been made publicly available by

home **cultures** **media & methods** **depositing strains** **new accessions** **news** **culturecollection**

strain information: 211/11b

Division/Phylum: Chlorophyta Class: Trebouxiophyceae

Chlorella vulgaris Beijerinck (1890)
CCAP 211/11B
Isolator: Beijerinck (1892)
Origin: Freshwater, Delft, Holland
Culture: Medium (1)EG-JM; Axenic; sub + cryo; liquid;(2)3N-BBM+V; Axenic; sub; agar slope
Other: SAG 211-11b; UTEX 259; IAM C-207; UTCC 111; SAG 211-11b; TYPE CULTURE; Images 1-3 by Tatjana Dajenkova, recommended in ecotoxicity testing directive 87/302/EEC;

Available Images (3): 1 2 3

Algaebase record for *Chlorella vulgaris*

References [52]
Davoren M & Fogarty AM (2004) A test battery for the ecotoxicological evaluation of the agri-chemical Environ. Ecotoxicology and Environmental Safety 59: 116-122. DOI: 10.1016/j.ecoenv.2004.01.001
González-Barreiro O, Rioboo C, Herrero C & Cid A (2005) Removal of triazine herbicides from freshwater systems using photosynthetic microorganisms. Environmental Pollution 144: 266-271. DOI: 10.1016/j.envpol.2005.12.014
Janssen M, Patino R & Von Stockar U (2005) Application of bench-scale biocalorimetry to photoautotrophic cultures. Thermochimica Acta 435(1): 18-27. DOI: 10.1016/j.tca.2005.04.024
Müller J, Friedl T, Hepperle D, Lorenz M & Day JG (2005) Distinction between multiple isolates of *Chlorella vulgaris* (Chlorophyta, Trebouxiophyceae) and testing for conspecificity using Amplified Fragment Length Polymorphism and its rDNA sequences. Journal of Phycology 41(6): 1236-1247. DOI: 10.1111/j.1529-8817.2005.00134.x
Jolley E & Smith DC (1980) The green hydra symbiosis. II. The biology of the

EMBL Links
X13688 (18S)
AY876300 (SSU (18S))
AY591507 (SSU (18S))
LSU (28S)

algaeBASE
content about team links contact search donate
127,449 species and subspecific names are in the database, 11,203 images, 45,761 bibliographic items, 176,777 distributional records.

Chlorella vulgaris Beijerinck

Publication details
Chlorella vulgaris Beijerinck 1890: 756, fig. 7:2 a-d
Original publication: Beijerinck, H.W. (1890). Cultureversuche mit Zoonkorden, Lithoenagenden und anderen niederen Algen. Botanische Zeitung 47: 725-729, 741-754, 757-768, 781-785.
Type species
This is the type species (lectotype) of the genus *Chlorella*.
Status of name
This name is of an entity that is currently accepted taxonomically.
Origin of species name
Adjective (Latin), common, general, ordinary (Strom 1973).
Heterotypic Synonym(s)
Chlorella pyrenoidosa var. *duplex* (Kützinger) West
Neurococcus bejerinckii Arsten 1892
Chlorella pyrenoidosa H.Chr. 1903
Chlorella communis Arsten 1904
Chlorella vulgaris var. *viridis* Choudh 1913
Chlorella terrella Golebach 1936
Chlorella candida Shihira & Krauss 1965
General environment
This is a freshwater species.

Journal of Phycology
AN INTERNATIONAL JOURNAL OF ALGAL RESEARCH

DISTINCTION BETWEEN MULTIPLE ISOLATES OF *CHLORELLA VULGARIS* (CHLOROPHYTA, TREBOUXIOPHYCEAE) AND TESTING FOR CONSPECIFICITY USING AMPLIFIED FRAGMENT LENGTH POLYMORPHISM AND ITS rDNA SEQUENCES[†]

Julia Müller¹, Thomas Friedl^{1,2}, Dominik Hepperle¹, Maika Lorenz¹, John G. Day²
Article first published online: 16 NOV 2005
DOI: 10.1111/j.1529-8817.2005.00134.x

Journal of Phycology
Volume 41, Issue 6, pages 1236–1247, December 2005

Additional Information (Show All)
How to Cite | Author Information | Publication History
[†]Received 16 March 2005; Accepted 18 August 2005.

Abstract | Article | References | Cited By
View Full Article (HTML) | Get PDF (307K)

Keywords:
AFLP, *Chlorella vulgaris*, cryopreservation, cryptic species, culture collections, genotyping, ITS, phylogenetic analyses, rDNA

Multiple strains of individual algal species are available from public culture collections, often with the same isolate being maintained in parallel at a number of collections under different culture regimes. To unravel genomic variation and to identify unique genotypes among such multiple strains, two approaches were used on a sample of 29 strains of *Chlorella vulgaris* Beijerinck, an alga of great value for applied research, from five culture collections. With the exception of two strains, internal transcribed spacer rDNA sequence data substantiated conspecificity of the studied strains and only minor sequence differences with the authentic 'Beijerinck isolate' were observed. Amplified fragment length polymorphism (AFLP) detected considerable genomic variation when rDNA sequences were identical. Band detection and the construction of a binary matrix from AFLP patterns for phylogenetic analyses were fully automated, but comparison of similar patterns still required manual refinement. The AFLPs distinguished 11 unique genotypes and provided robust support for the presence of five cryptic species. This finding advocates the need to carefully record which strain has been used in any experiment or in applied research, because genomic variation may also correspond to differences in physiological/biochemical properties. No genomic differences could be detected between duplicate strains of the same isolate that were maintained by continuous subculturing over many decades or within those stored at ultralow temperatures.

View Full Article (HTML) | Get PDF (307K)

ENA
Text search | Sequence search
Enter or paste text or ENA accession number
Upload file of ENA accessions
[Choose File] No file chosen [Go] [Search]
EMBL Bank: AF591506.1: *Chlorella vulgaris* strain CCAP 211/11B 18S ribosomal RNA gene, partial sequence
View: 1237 1238 1239 1240 1241 1242 1243 1244 1245 1246 1247 1248 1249 1250 1251 1252 1253 1254 1255 1256 1257 1258 1259 1260 1261 1262 1263 1264 1265 1266 1267 1268 1269 1270 1271 1272 1273 1274 1275 1276 1277 1278 1279 1280 1281 1282 1283 1284 1285 1286 1287 1288 1289 1290 1291 1292 1293 1294 1295 1296 1297 1298 1299 1300 1301 1302 1303 1304 1305 1306 1307 1308 1309 1310 1311 1312 1313 1314 1315 1316 1317 1318 1319 1320 1321 1322 1323 1324 1325 1326 1327 1328 1329 1330 1331 1332 1333 1334 1335 1336 1337 1338 1339 1340 1341 1342 1343 1344 1345 1346 1347 1348 1349 1350 1351 1352 1353 1354 1355 1356 1357 1358 1359 1360 1361 1362 1363 1364 1365 1366 1367 1368 1369 1370 1371 1372 1373 1374 1375 1376 1377 1378 1379 1380 1381 1382 1383 1384 1385 1386 1387 1388 1389 1390 1391 1392 1393 1394 1395 1396 1397 1398 1399 1400 1401 1402 1403 1404 1405 1406 1407 1408 1409 1410 1411 1412 1413 1414 1415 1416 1417 1418 1419 1420 1421 1422 1423 1424 1425 1426 1427 1428 1429 1430 1431 1432 1433 1434 1435 1436 1437 1438 1439 1440 1441 1442 1443 1444 1445 1446 1447 1448 1449 1450 1451 1452 1453 1454 1455 1456 1457 1458 1459 1460 1461 1462 1463 1464 1465 1466 1467 1468 1469 1470 1471 1472 1473 1474 1475 1476 1477 1478 1479 1480 1481 1482 1483 1484 1485 1486 1487 1488 1489 1490 1491 1492 1493 1494 1495 1496 1497 1498 1499 1500 1501 1502 1503 1504 1505 1506 1507 1508 1509 1510 1511 1512 1513 1514 1515 1516 1517 1518 1519 1520 1521 1522 1523 1524 1525 1526 1527 1528 1529 1530 1531 1532 1533 1534 1535 1536 1537 1538 1539 1540 1541 1542 1543 1544 1545 1546 1547 1548 1549 1550 1551 1552 1553 1554 1555 1556 1557 1558 1559 1560 1561 1562 1563 1564 1565 1566 1567 1568 1569 1570 1571 1572 1573 1574 1575 1576 1577 1578 1579 1580 1581 1582 1583 1584 1585 1586 1587 1588 1589 1590 1591 1592 1593 1594 1595 1596 1597 1598 1599 1600 1601 1602 1603 1604 1605 1606 1607 1608 1609 1610 1611 1612 1613 1614 1615 1616 1617 1618 1619 1620 1621 1622 1623 1624 1625 1626 1627 1628 1629 1630 1631 1632 1633 1634 1635 1636 1637 1638 1639 1640 1641 1642 1643 1644 1645 1646 1647 1648 1649 1650 1651 1652 1653 1654 1655 1656 1657 1658 1659 1660 1661 1662 1663 1664 1665 1666 1667 1668 1669 1670 1671 1672 1673 1674 1675 1676 1677 1678 1679 1680 1681 1682 1683 1684 1685 1686 1687 1688 1689 1690 1691 1692 1693 1694 1695 1696 1697 1698 1699 1700 1701 1702 1703 1704 1705 1706 1707 1708 1709 1710 1711 1712 1713 1714 1715 1716 1717 1718 1719 1720 1721 1722 1723 1724 1725 1726 1727 1728 1729 1730 1731 1732 1733 1734 1735 1736 1737 1738 1739 1740 1741 1742 1743 1744 1745 1746 1747 1748 1749 1750 1751 1752 1753 1754 1755 1756 1757 1758 1759 1760 1761 1762 1763 1764 1765 1766 1767 1768 1769 1770 1771 1772 1773 1774 1775 1776 1777 1778 1779 1780 1781 1782 1783 1784 1785 1786 1787 1788 1789 1790 1791 1792 1793 1794 1795 1796 1797 1798 1799 1800 1801 1802 1803 1804 1805 1806 1807 1808 1809 1810 1811 1812 1813 1814 1815 1816 1817 1818 1819 1820 1821 1822 1823 1824 1825 1826 1827 1828 1829 1830 1831 1832 1833 1834 1835 1836 1837 1838 1839 1840 1841 1842 1843 1844 1845 1846 1847 1848 1849 1850 1851 1852 1853 1854 1855 1856 1857 1858 1859 1860 1861 1862 1863 1864 1865 1866 1867 1868 1869 1870 1871 1872 1873 1874 1875 1876 1877 1878 1879 1880 1881 1882 1883 1884 1885 1886 1887 1888 1889 1890 1891 1892 1893 1894 1895 1896 1897 1898 1899 1900 1901 1902 1903 1904 1905 1906 1907 1908 1909 1910 1911 1912 1913 1914 1915 1916 1917 1918 1919 1920 1921 1922 1923 1924 1925 1926 1927 1928 1929 1930 1931 1932 1933 1934 1935 1936 1937 1938 1939 1940 1941 1942 1943 1944 1945 1946 1947 1948 1949 1950 1951 1952 1953 1954 1955 1956 1957 1958 1959 1960 1961 1962 1963 1964 1965 1966 1967 1968 1969 1970 1971 1972 1973 1974 1975 1976 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030 2031 2032 2033 2034 2035 2036 2037 2038 2039 2040 2041 2042 2043 2044 2045 2046 2047 2048 2049 2050 2051 2052 2053 2054 2055 2056 2057 2058 2059 2060 2061 2062 2063 2064 2065 2066 2067 2068 2069 2070 2071 2072 2073 2074 2075 2076 2077 2078 2079 2080 2081 2082 2083 2084 2085 2086 2087 2088 2089 2090 2091 2092 2093 2094 2095 2096 2097 2098 2099 2100 2101 2102 2103 2104 2105 2106 2107 2108 2109 2110 2111 2112 2113 2114 2115 2116 2117 2118 2119 2120 2121 2122 2123 2124 2125 2126 2127 2128 2129 2130 2131 2132 2133 2134 2135 2136 2137 2138 2139 2140 2141 2142 2143 2144 2145 2146 2147 2148 2149 2150 2151 2152 2153 2154 2155 2156 2157 2158 2159 2160 2161 2162 2163 2164 2165 2166 2167 2168 2169 2170 2171 2172 2173 2174 2175 2176 2177 2178 2179 2180 2181 2182 2183 2184 2185 2186 2187 2188 2189 2190 2191 2192 2193 2194 2195 2196 2197 2198 2199 2200 2201 2202 2203 2204 2205 2206 2207 2208 2209 2210 2211 2212 2213 2214 2215 2216 2217 2218 2219 2220 2221 2222 2223 2224 2225 2226 2227 2228 2229 2230 2231 2232 2233 2234 2235 2236 2237 2238 2239 2240 2241 2242 2243 2244 2245 2246 2247 2248 2249 2250 2251 2252 2253 2254 2255 2256 2257 2258 2259 2260 2261 2262 2263 2264 2265 2266 2267 2268 2269 2270 2271 2272 2273 2274 2275 2276 2277 2278 2279 2280 2281 2282 2283 2284 2285 2286 2287 2288 2289 2290 2291 2292 2293 2294 2295 2296 2297 2298 2299 2300 2301 2302 2303 2304 2305 2306 2307 2308 2309 2310 2311 2312 2313 2314 2315 2316 2317 2318 2319 2320 2321 2322 2323 2324 2325 2326 2327 2328 2329 2330 2331 2332 2333 2334 2335 2336 2337 2338 2339 2340 2341 2342 2343 2344 2345 2346 2347 2348 2349 2350 2351 2352 2353 2354 2355 2356 2357 2358 2359 2360 2361 2362 2363 2364 2365 2366 2367 2368 2369 2370 2371 2372 2373 2374 2375 2376 2377 2378 2379 2380 2381 2382 2383 2384 2385 2386 2387 2388 2389 2390 2391 2392 2393 2394 2395 2396 2397 2398 2399 2400 2401 2402 2403 2404 2405 2406 2407 2408 2409 2410 2411 2412 2413 2414 2415 2416 2417 2418 2419 2420 2421 2422 2423 2424 2425 2426 2427 2428 2429 2430 2431 2432 2433 2434 2435 2436 2437 2438 2439 2440 2441 2442 2443 2444 2445 2446 2447 2448 2449 2450 2451 2452 2453 2454 2455 2456 2457 2458 2459 2460 2461 2462 2463 2464 2465 2466 2467 2468 2469 2470 2471 2472 2473 2474 2475 2476 2477 2478 2479 2480 2481 2482 2483 2484 2485 2486 2487 2488 2489 2490 2491 2492 2493 2494 2495 2496 2497 2498 2499 2500 2501 2502 2503 2504 2505 2506 2507 2508 2509 2510 2511 2512 2513 2514 2515 2516 2517 2518 2519 2520 2521 2522 2523 2524 2525 2526 2527 2528 2529 2530 2531 2532 2533 2534 2535 2536 2537 2538 2539 2540 2541 2542 2543 2544 2545 2546 2547 2548 2549 2550 2551 2552 2553 2554 2555 2556 2557 2558 2559 2560 2561 2562 2563 2564 2565 2566 2567 2568 2569 2570 2571 2572 2573 2574 2575 2576 2577 2578 2579 2580 2581 2582 2583 2584 2585 2586 2587 2588 2589 2590 2591 2592 2593 2594 2595 2596 2597 2598 2599 2600 2601 2602 2603 2604 2605 2606 2607 2608 2609 2610 2611 2612 2613 2614 2615 2616 2617 2618 2619 2620 2621 2622 2623 2624 2625 2626 2627 2628 2629 2630 2631 2632 2633 2634 2635 2636 2637 2638 2639 2640 2641 2642 2643 2644 2645 2646 2647 2648 2649 2650 2651 2652 2653 2654 2655 2656 2657 2658 2659 2660 2661 2662 2663 2664 2665 2666 2667 2668 2669 2670 2671 2672 2673 2674 2675 2676 2677 2678 2679 2680 2681 2682 2683 2684 2685 2686 2687 2688 2689 2690 2691 2692 2693 2694 2695 2696 2697 2698 2699 2700 2701 2702 2703 2704 2705 2706 2707 2708 2709 2710 2711 2712 2713 2714 2715 2716 2717 2718 2719 2720 2721 2722 2723 2724 2725 2726 2727 2728 2729 2730 2731 2732 2733 2734 2735 2736 2737 2738 2739 2740 2741 2742 2743 2744 2745 2746 2747 2748 2749 2750 2751 2752 2753 2754 2755 2756 2757 2758 2759 2760 2761 2762 2763 2764 2765 2766 2767 2768 2769 2770 2771 2772 2773 2774 2775 2776 2777 2778 2779 2780 2781 2782 2783 2784 2785 2786 2787 2788 2789 2790 2791 2792 2793 2794 2795 2796 2797 2798 2799 2800 2801 2802 2803 2804 2805 2806 2807 2808 2809 2810 2811 2812 2813 2814 2815 2816 2817 2818 2819 2820 2821 2822 2823 2824 2825 2826 2827 2828 2829 2830 2831 2832 2833 2834 2835 2836 2837 2838 2839 2840 2841 2842 2843 2844 2845 2846 2847 2848 2849 2850 2851 2852 2853 2854 2855 2856 2857 2858 2859 2860 2861 2862 2863 2864 2865 2866 2867 2868 2869 2870 2871 2872 2873 2874 2875 2876 2877 2878 2879 2880 2881 2882 2883 2884 2885 2886 2887 2888 2889 2890 2891 2892 2893 2894 2895 2896 2897 2898 2899 2900 2901 2902 2903 2904 2905 2906 2907 2908 2909 2910 2911 2912 2913 2914 2915 2916 2917 2918 2919 2920 2921 2922 2923 2924 2925 2926 2927 2928 2929 2930 2931 2932 2933 2934 2935 2936 2937 2938 2939 2940 2941 2942 2943 2944 2945 2946 2947 2948 2949 2950 2951 2952 2953 2954 2955 2956 2957 2958 2959 2960 2961 2962 2963 2964 2965 2966 2967 2968 2969 2970 2971 2972 2973 2974 2975 2976 2977 2978 2979 2980 2981 2982 2983 2984 2985 2986 2987 2988 2989 2990 2991 2992 2993 2994 2995 2996 2997 2998 2999 3000 3001 3002 3003 3004 3005 3006 3007 3008 3009 3010 3011 3012 3013 3014 3015 3016 3017 3018 3019 3020 3021 3022 3023 3024 3025 3026 3027 3028 3029 3030 3031 3032 3033 3034 3035 3036 3037 3038 3039 3040 3041 3042 3043 3044 3045 3046 3047 3048 3049 3050 3051 3052 3053 3054 3055 3056 3057 3058 3059 3060 3061 3062 3063 3064 3065 3066 3067 3068 3069 3070 3071 3072 3073 3074 3075 3076 3077 3078 3079 3080 3081 3082 3083 3084 3085 3086 3087 3088 3089 3090 3091 3092 3093 3094 3095 3096 3097 3098 3099 3100 3101 3102 3103 3104 3105 3106 3107 3108 3109 3110 3111 3112 3113 3114 3115 3116 3117 3118 3119 3120 3121 3122 3123 3124 3125 3126 3127 3128 3129 3130 3131 3132 3133 3134 3135 3136 3137 3138 3139 3140 3141 3142 3143 3144 3145 3146 3147 3148 3149 3150 3151 3152 3153 3154 3155 3156 3157 3158 3159 3160 3161 3162 3163 3164 3165 3166 3167 3168 3169 3170 3171 3172 3173 3174 3175 3176 3177 3178 3179 3180 3181 3182 3183 3184 3185 3186 3187 3188 3189 3190 3191 3192 3193 3194 3195 3196 3197 3198 3199 3200 3201 3202 3203 3204 3205 3206 3207 3208 3209 3210 3211 3212 3213 3214 3215 3216 3217 3218 3219 3220 3221 3222 3223 3224 3225 3226 3227 3228 3229 3230 3231 3232 3233

other laboratories deposited in the ENA. Over 1350 CCAP strain page records are now hyperlinked with at least one ENA accession number (Fig. 6). Our primary objective is to obtain an accurate and reliable DNA identification of strain holdings, typically using ribosomal operon sequences, in order to guide our users in their selection of biological material. This undertaking also provides the foundations for a state-of-the-art molecular-enabled taxonomic revision of the CCAP's holdings. It is envisaged that by employing both molecular data, as well as conventional phenotypic information, definitive names can be confirmed or revised for individual strains in the collection; this in turn will result in enhancement of quality assurance within the collection.

Taxonomic information: AlgaeBase

AlgaeBase (<http://www.algaebase.org>) has grown over the last 18 years into the most comprehensive and up-to-date online resource on algal taxonomy available. It currently includes > 133 000 species and infraspecific names, of which 36 000 are the names of taxonomically accepted species, > 17 000 images, > 50 000 bibliographic items and > 220 000 distributional records (Guiry & Guiry, 2013). Its content is very relevant to CCAP, as it includes freshwater, terrestrial and marine microalgae and macroalgae, and prokaryotic cyanobacteria, together with achlorophyllous protists including chromists, dinoflagellates and euglenoid flagellates.

AlgaeBase entries for algae are linked to a number of biological resource centres and repositories of cultures, including CCAP, which has > 2400 strains linked to their taxonomic data on AlgaeBase. We have introduced reciprocal systematic links from CCAP strain records to AlgaeBase (Fig. 4), in order to provide users of the KnowledgeBase with additional expert taxonomic information. It should be noted, however, that this information is given for indicative purposes only. Indeed, AlgaeBase records are linked to the scientific name assigned to a given CCAP strain, and therefore are subject to the accuracy of its current taxonomic assignment. As explained above, the delineation and taxonomic placement of many protistan taxa remains imperfect; hence, with the exception of ex-type material (commonly described as authentic strains in algal collections), the identification accompanying CCAP strain accessions is a useful guide, but does not constitute a definitive taxonomic reference.

Future directions

Our objective is to add further value to the information available on the CCAP biological material by progressing its integration into a manually curated platform. We hope that our efforts to develop a user-friendly interface will support the breadth of activities of our users, including

systematic research, biotechnological exploitation and teaching. The CCAP KnowledgeBase is continuously updated, and we therefore especially welcome feedback from our users, including taxonomic updates, pictures or publications generated with material obtained from the CCAP. Arguably, comprehensive 'Knowledge-bases' are of increasing value to the scientific community and curated links between live material and bioinformatics data are of vital importance, particularly in the light of concerns over errors (taxonomic and others) in data published in the major depositories of molecular data (Bridge *et al.*, 2003; Tindall, 2007). Usage statistics indicate a steady increase in visitors interrogating the CCAP KnowledgeBase having increased from ~26 400 in 2008 to ~35 400 in 2012 (data from Google Analytics; <http://www.google.com/analytics>). Further developments currently under consideration include additional links to other data types, such as metabolomics or chemical fingerprinting. We also aim to further underpin the exponential development of metagenomics and genomics of unconventional and novel model organisms; in particular with the cross-referencing of environmental molecular data to specified live culturable material (del Campo & Massana, 2011). Our objective is to ensure that the CCAP KnowledgeBase is 'fit for purpose' for 21st century science with improved searchability, better connectivity to other relevant resources and all data are underpinned by live cultures available to the scientific community worldwide.

Acknowledgements

We are grateful to all former and current SAMS staff and visiting scientists who contributed to this project and those who gave us feedback on the KnowledgeBase. Author contributions: writing (CMMG, JGD, FCK, MDG, KG); data generation and knowledgebase development (CMMG, SH, UEMA-D, DB, CNC, AC, RGD, JF, SG, CR-M, AV, MDG, KG, JGD), CCAP KnowledgeBase concept (FCK, CNC, SG, MB, JGD). This work was supported by the UK Natural Environment Research Council (Oceans 2025 programme NF3 – CCAP), NERC NC funding for the CCAP; and the European Commission ASSEMBLE Integrated Infrastructures Initiative (grant agreement no. 227799). A large fraction of the sequence data in CCAP were generated by the GenePool Natural Environment Research Council Biomolecular Analysis Facility at the University of Edinburgh (allocation no. MGF154).

References

- ADL, S.M., SIMPSON, A.G.B., LANE, C.E., LUKES, J., BASS, D., BOWSER, S.S., BROWN, M.W., BURKI, F., DUNTHORN, M., HAMPL, V., HEISS, A., HOPPENRATH, M., LARA, E., LE GALL, L., LYNN, D.H., McMANUS, H., MITCHELL, E.A.D., MOZLEY-STANRIDGE, S.E., PARFREY, L.W., PAWLOWSKI, J., RUECKERT,

- S., SHADWICK, L., SCHOCH, C.L., SMIRNOV, A. & SPIEGEL, F.W. 2012. The revised classification of eukaryotes. *Journal of Eukaryotic Microbiology* **59**, 429–514.
- ARCHIBALD, J.M. 2011. Newly identified and diverse plastid-bearing branch on the eukaryotic tree of life. *Proceedings of the National Academy of Sciences USA* **108**, 1496–1500.
- BALDAUF, S.L. 2008. An overview of the phylogeny and diversity of eukaryotes. *Journal of Systematics and Evolution* **46**, 263–273.
- BERNINGER, U.G., FINLAY, B.J. & KUUPPOLEINIKKI, P. 1991. Protozoan control of bacterial abundances in fresh-water. *Limnology and Oceanography* **36**, 139–147.
- BRIDGE, P.D., ROBERTS, P.J., SPOONER, B.M. & PANCHAL, G. 2003. On the reliability of published DNA sequences. *New Phytologist* **160**, 43–48.
- CARON, D.A. & HUTCHINS, D.A. 2013. The effects of changing climate on microzooplankton grazing and community structure: drivers, predictions and knowledge gaps. *Journal of Plankton Research* **35**, 235–252.
- CHAMBOUVET, A., MORIN, P., MARIE, D. & GUILLLOU, L. 2008. Control of toxic marine dinoflagellate blooms by serial parasitic killers. *Science* **322**:1254–1257.
- CHAVEZ, F.P., MESSIE, M. & PENNINGTON, J.T. 2011. Marine primary production in relation to climate variability and change. *Annual Review of Marine Science* **3**, 227–260.
- CHRISTAKI, U., JACQUET, S., DOLAN, J.R., VAULOT, D. & RASSOUL-ZADEGAN, F. 1999. Growth and grazing on *Prochlorococcus* and *Synechococcus* by two marine ciliates. *Limnology and Oceanography* **44**: 52–61.
- DAY, J.G. & STACEY, G.N. 2008. Biobanking. *Molecular Biotechnology* **40**, 202–213.
- DAY, J.G., PRÖSCHOLD, T., FRIEDL, T., LORENZ, M. & SILVA, P.C. 2010. Conservation of microalgal type material: approaches needed for 21st century science. *Taxon* **59**, 3–6.
- DEGERLUND, M., HUSEBY, S., ZINGONE, A., SARNO, D. & LANDFALD, B. 2012. Functional diversity in cryptic species of *Chaetoceros socialis* Lauder (Bacillariophyceae). *Journal of Plankton Research* **34**, 416–431.
- DEL CAMPO, J. & MASSANA, R. 2011. Emerging diversity within Chrysophytes, Choanoflagellates and Bicosoecids based on molecular surveys. *Protist*, **162**, 435–448.
- FALKOWSKI, P.G., KATZ, M.E., KNOLL, A.H., QUIGG, A., RAVEN, J.A., SCHOFIELD, O. & TAYLOR, F.J.R. 2004. The evolution of modern eukaryotic phytoplankton. *Science* **305**, 354–360.
- GACHON, C.M.M., DAY, J.G., CAMPBELL, C.N., PRÖSCHOLD, T., SAXON, R.J. & KÜPPER, F.C. 2007. The Culture Collection of Algae and Protozoa (CCAP): a biological resource for protistan genomics. *Gene* **406**, 51–57.
- GACHON, C.M.M., SIME-NGANDO, T., STRITTMATTER, M., CHAMBOUVET, A. & KIM, G.H. 2010. Algal diseases: spotlight on a black box. *Trends in Plant Science* **15**, 633–640.
- GLÜCKSMAN, E., SNELL, E.A., BERNEY, C., CHAO, E.E., BASS, D. & CAVALIER-SMITH, T. 2011. The novel marine gliding zooflagellate genus *Mantamonas* (Mantamonadida ord. n.: Apusozoa). *Protist* **162**, 207–221.
- GUIRY, M.D. & GUIRY, G.M. 2013. *AlgaeBase*. World-wide electronic publication, National University of Ireland, Galway. <http://www.algaebase.org>; accessed 7 May 2013.
- HARTMANN, M., GROB, C., TARRAN, G.A., MARTIN, A.P., BURKILL, P.H., SCANLAN, D.J. & ZUBKOV, M.V. 2012. Mixotrophic basis of Atlantic oligotrophic ecosystems. *Proceedings of the National Academy of Sciences USA* **109**, 5756–5760.
- KEELING, P.J., BURGER, G., DURNFORD, D.G., LANG, B.F., LEE, R.W., PEARLMAN, R.E., ROGER, A.J. & GRAY, M.W. 2005. The tree of eukaryotes. *Trends in Ecology and Evolution* **20**, 670–676.
- KIM, E., HARRISON, J.W., SUDEK, S., JONES, M.D.M., WILCOX, H. M., RICHARDS, T.A., WORDEN, A.Z. & ARCHIBALD, J.M. 2011. Newly identified and diverse plastid-bearing branch on the eukaryotic tree of life. *Proceedings of the National Academy of Sciences USA* **108**, 1496–1500.
- KIMURA, H., SATO, M., SUGIYAMA, C. & NAGANUMA, T. 2001. Coupling of thraustochytrids and POM, and of bacterio- and phytoplankton in a semi-enclosed coastal area: implication for different substrate preference by the planktonic decomposers. *Aquatic Microbial Ecology* **25**, 293–300.
- KÜPPER, F.C., FEITERS, M.C., OLOFSSON, B., KAIHO, T., YANAGIDA, S., ZIMMERMANN, M.B., CARPENTER, L.J., LUTHER III, G.W., LU, Z., JONSSON, M. & KLOO, L. 2011. Commemorating two centuries of iodine research: an interdisciplinary overview of current research. *Angewandte Chemie – International Edition* **50**, 11598–11620.
- MASSANA, R., TERRADO, R., FORN, I., LOVEJOY, C. & PEDROS-ALIO, C. 2006. Distribution and abundance of uncultured heterotrophic flagellates in the world oceans. *Environmental Microbiology* **8**, 1515–1522.
- MILLEDEGE, J.J. 2011. Commercial application of microalgae other than as biofuels: a brief review. *Reviews in Environmental Science and Biotechnology* **10**, 31–41.
- MORRISSEY, J. & BOWLER, C. 2012. Iron utilization in marine cyanobacteria and eukaryotic algae. *Frontiers in Microbiology* **3**, 43.
- MÜLLER J., FRIEDL T., HEPPERLE, D., LORENZ, M. & DAY, J.G. 2005. Distinction of isolates among multiple strains of *Chlorella vulgaris* (Chlorophyta, Trebouxiophyceae) and testing conspecificity with Amplified Fragment Length Polymorphism and ITS rDNA sequences. *Journal of Phycology* **41**, 1236–1247.
- POPPER, Z.A., MICHEL, G., HERVÉ, C., DOMOZYCH, D.S., WILLATS, W.G.T., TUOHY, M.G., KLOAREG, B. & STENGEL, D.B. 2011. Evolution and diversity of plant cell walls: from algae to flowering plants. *Annual Review in Plant Biology* **62**, 567–590.
- PRÖSCHOLD, T., MARIN, B., SCHLÖSSER, U.G. & MELKONIAN, M. 2001. Molecular phylogeny and taxonomic revision of *Chlamydomonas* (Chlorophyta). I. Emendation of *Chlamydomonas* Ehrenberg and *Chloromonas* Gobi, and description of *Oogamochlamys* gen. nov and *Lobochlamys* gen. nov. *Protist* **152**, 265–300.
- RYBALK, N., WOLF, M., ANDERSEN, R.A. & FRIEDL, T. 2013. Congruence of chloroplast- and nuclear-encoded DNA sequence variations used to assess species boundaries in the soil microalga *Heterococcus* (Stramenopiles, Xanthophyceae). *BioMed Central Evolutionary Biology* **13**, Art. No. 39.
- SMIRNOV, A.V., CHAO, E., NASSONOVA, E.S. & CAVALIER-SMITH, T. 2011. A revised classification of naked lobose amoebae (Amoebozoa: Lobosa). *Protist* **162**, 545–570.
- STEINER, C.F., LONG, Z.T., KRUMINS, J.A. & MORIN, P.J. 2006. Population and community resilience in multitrophic communities. *Ecology* **87**, 996–1007.
- TINDALL, B.J. 2007. Vacuum drying and cryopreservation of prokaryotes. *Methods in Molecular Biology* **368**, 73–98.
- TOMPKINS, J., DEVILLE, M.M., DAY, J.G. & TURNER, M.F. 1995. *Culture Collection of Algae and Protozoa Catalogue of strains*, 6th Edition. Culture Collection of Algae and Protozoa, Ambleside.
- WILLIAMS, P.J.L. & LAURENS, L.M.L. 2010. Microalgae as biodiesel and biomass feedstocks: review and analysis of the biochemistry, energetics and economics. *Energy and Environmental Science* **3**, 554–590.