

# ИДЕНТИФИКАЦИЯ *IN SILICO* ПОСЛЕДОВАТЕЛЬНОСТЕЙ, КОДИРУЮЩИХ БЕЛКИ СЕМЕЙСТВА CLC, В СОБРАННЫХ *DE NOVO* ТРАНСКРИПТОМАХ МОРСКОЙ МИКРОВОДОРОСЛИ *DUNALIELLA TERTIOLECTA*

Л.Г. Попова<sup>1</sup>, А.В. Шувалов<sup>1</sup>, А.А. Юрченко<sup>2</sup>, Д.Е. Храмов<sup>1,3</sup>, Д.А. Маталин<sup>1</sup>, Л.А. Халилова<sup>1</sup>, Ю.В. Орлова<sup>1</sup>, Ю.В. Балнокин<sup>1,3</sup>

<sup>1</sup>Федеральное государственное бюджетное учреждение науки Институт физиологии растений им. К.А. Тимирязева Российской академии наук, Москва, Россия, lora\_gp@mail.ru

<sup>2</sup>Федеральное государственное бюджетное образовательное учреждение высшего образования «Санкт-Петербургский государственный университет», Центр геномной биоинформатики им. Ф.Г. Добржанского, Санкт-Петербург, Россия

<sup>3</sup>Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский государственный университет имени М.В. Ломоносова», Москва, Россия

**Аннотация.** Осуществлена *de novo* сборка транскриптомов микроводоросли *Dunaliella tertiolecta* на основе представленных в базе Sequence Read Archive (NCBI) библиотек коротких прочтений РНК из этой водоросли. В собранных транскриптомах найдены *in silico* последовательности, кодирующие Cl<sup>-</sup>-транспортирующие белки семейства CLC: DtCLC1 (857 а.о.), DtCLC2 (809 а.о.), DtCLC3 (768 а.о.). Эти белки сходны с белками CLC высших растений и животных (прежде всего рыб и рептилий), осуществляющих Cl<sup>-</sup>/H<sup>+</sup> обмен.

**Ключевые слова:** *de novo* сборка транскриптома, белки CLC, морские микроводоросли, *Dunaliella tertiolecta*

**DOI:** 10.31255/978-5-94797-319-8-978-982

**Введение.** В основе солеустойчивости высших растений лежат механизмы, определяющие способность к ионному гомеостатированию цитоплазмы [Munns, Tester, 2008]. Среди них важную роль играют механизмы транспорта Na<sup>+</sup> и Cl<sup>-</sup>. Регуляция концентраций этих ионов в цитоплазме осуществляется как на уровне плазматической мембраны, которая является основным барьером, отделяющим цитоплазму клетки от внешней среды, так и на уровне эндомембран клетки. И если в настоящее время о механизмах транспорта Na<sup>+</sup> у растений известно достаточно много [Munns, Tester, 2008], механизмы транспорта Cl<sup>-</sup> изучены недостаточно, несмотря на то, что этот ион является одним из доминирующих при почвенном засолении [Teakle, Tyerman, 2010].

Существенный вклад в регуляцию внутриклеточных концентраций Cl<sup>-</sup> играют белки семейства CLC (Chloride Channel). Эти мембранные белки широко распространены в живой природе и найдены у представителей всех царств [Jentsch, 2008]. Семейство CLC включает в себя анионные каналы и анион/протонные антипортеры, в частности, Cl<sup>-</sup>/H<sup>+</sup>-антипортеры, транспорт Cl<sup>-</sup> которыми через мембраны осуществляется за счет протон-движущей силы. У эукариот известные Cl<sup>-</sup>/H<sup>+</sup>-антипортеры локализованы в мембранах внутриклеточных органелл, тогда как у прокариот – в плазматической мембране. В животных клетках Cl<sup>-</sup>/H<sup>+</sup>-антипортеры осуществляют Cl<sup>-</sup>/H<sup>+</sup>-обмен между цитозолем и внутриклеточными компартментами, у бактерий – между цитозолем и наружной средой. В растительных клетках Cl<sup>-</sup>/H<sup>+</sup>-антипортеры также осуществляют экспорт Cl<sup>-</sup> из цитоплазмы во внутриклеточные органеллы. У модельного растения *Arabidopsis thaliana* обнаружено семь представителей семейства CLC, которые локализованы в различных внутриклеточных мембранах: AtCLCa, AtCLCb, AtCLCc и AtCLCg – в тонопласте, AtCLCd и AtCLCf – в аппарате Гольджи, AtCLCe – в тилакоидах.

Галотолерантные водоросли, способные расти в соленых средах – морской воде или соленых озерах, вода в которых может содержать NaCl в высоких концентрациях, вплоть до насыщающих, – можно рассматривать как модельные объекты для исследования механизмов солеустойчивости на клеточном уровне. У галотолерантных микроводорослей плазматическая мембрана клетки непосредственно контактирует с наружной средой. Перемещение ионов через плазмалемму посредством механизмов, локализованных в этой мембране, играет центральную роль в поддержании ионного, в т.ч. Cl<sup>-</sup>, гомеостаза в клетках морских водорослей. Известно, что содержание Cl<sup>-</sup> в клетках водорослей рода *Dunaliella*, как и содержание Na<sup>+</sup>, поддерживается на постоянно низком уровне в широком диапазоне наружных концентраций NaCl [Kirst, 1977; Балнокин и др., 1990]. Кроме того, транспорт Cl<sup>-</sup>, наряду с транспортом Na<sup>+</sup> и K<sup>+</sup>, важен для поддержания осмотического баланса в клетке. Следовательно, клетки морских водорослей должны обладать механизмами, отвечающими за регулирование потоков этого иона как между клеткой и наружной средой, так и во внутриклеточных компартментах. Анионные переносчики у водорослей играют фундаментальную роль в биологии клетки, однако мало известно о специфических функциях этих транспортеров, их локализации, молекулярной природе, включая белки семейства CLC (за исключением одного белка, относящегося к семейству CLC, – H<sup>+</sup>/Cl<sup>-</sup> антипортера термофильной водоросли *Cyanidioschyzon merolae* [Feng et al., 2010]).

Морские микроводоросли рода *Dunaliella* относятся к большому семейству зеленых одноклеточных микроводорослей Dunaliellaceae, представители которого повсеместно распространены в водах морей и соленых водоемов. Водоросли рода *Dunaliella* обладают высокой устойчивостью к засолению NaCl [Massjuk, 1973]. Механизмы Na<sup>+</sup> гомеостатирования у *Dunaliella* активно изучаются [Попова, Балнокин, 2013], тогда как механизмам Cl<sup>-</sup> гомеостатирования до настоящего времени внимания уделялось мало. Наше теоретическое исследование направлено на поиск белков семейства CLC у микроводорослей рода *Dunaliella* – эвригалинной микроводоросли *D. tertiolecta*, способной расти в широком диапазоне солености среды (0,07 М – 2 М NaCl) [Масюк, 1973]. Мы осуществили сборку *de novo* транскриптомов *D. tertiolecta* на основе нескольких массивов коротких прочтений РНК из этой водоросли (представлены в базе Sequence Read Archive, NCBI), и в собранных транскриптомах провели *in silico* поиск последовательностей, кодирующих белки, относящиеся к семейству CLC.

**Методы исследования.** Массивы коротких прочтений РНК для сборки транскриптомов *D. tertiolecta* были взяты из базы данных SRA (Sequence Read Archive, NCBI, [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)). Идентификаторы библиотек прочтений (код эксперимента/код архива): SRX047443/runSRR124253, SRX047444/runSRR124254, SRX549041/runSRR1294427, SRX554105/runSRR1300297, SRX554106/runSRR1300298. Качество извлечённых из библиотек SRA коротких прочтений РНК было проверено с помощью программы FastQC ([www.bioinformatics.babraham.ac.uk/projects/fastqc/](http://www.bioinformatics.babraham.ac.uk/projects/fastqc/)). Нуклеотиды низкого качества, остатки адаптеров Illumina и неопределённые нуклеотиды удалили с помощью программы Trimmomatic [Bolger et al., 2014]. Очищенные риды были собраны в контиги с помощью программы Trinity [Haas et al., 2013]. Оценку качества *de novo* сборки транскриптомов осуществляли, выравнивая прочтения обратно на контиги с помощью программы bowtie2 [Langmead, Salzberg, 2012]. Поиск интересных последовательностей в файлах с контигами осуществляли с помощью программы tblastn [Altschul, 1997]. Виртуальную трансляцию нуклеотидных последовательностей в аминокислотные осуществляли с помощью сервиса на сайте ExPASy ([web.expasy.org/translate/](http://web.expasy.org/translate/)). Размеры предсказанных белков определяли, исходя из присутствия в собранных контигах протяженных открытых рамок считывания.

Функции предсказанных белков анализировали с помощью ресурса InterPro (<http://www.ebi.ac.uk/interpro/>). Сравнительный анализ найденных последовательностей осуществляли также с помощью алгоритма blastx (BLAST, NCBI, <http://www.ncbi.nlm.nih.gov/>). Филогенетическое дерево построено с помощью программы proml, входящей в пакет программ Phylip (<http://evolution.genetics.washington.edu/phylip/>). Дерево визуализировали в программе DrawTree ([http://phylogeny.lirmm.fr/phylo.cgi/one\\_task.cgi?task\\_type=drawtree](http://phylogeny.lirmm.fr/phylo.cgi/one_task.cgi?task_type=drawtree)).

**Результаты и обсуждение.** Для реконструкции полных нуклеотидных последовательностей транскриптомов *D. tertiolecta* была осуществлена *de novo* сборка транскриптомов на основе представленных в базе SRA индивидуальных библиотек коротких прочтений РНК.

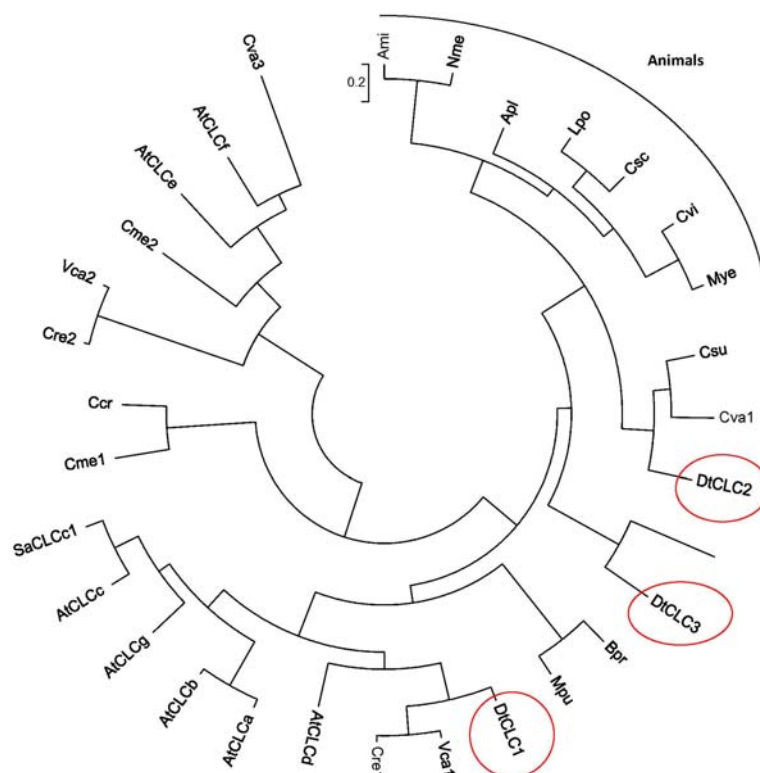


Рисунок. Кладограмма белков семейства CLC. Обозначения: Ami, *Alligator mississippiensis* (KYO26137.1); Apl, *Acanthaster planci* (XP\_022108376.1); Bpr, *Bathycoccus prasinus* (CCO18009.1); Ccr, *Chondrus crispus* (XP\_005711492.1); Cre1, *Chlamydomonas reinhardtii* (XP\_001689664.1); Cre2, *Chlamydomonas reinhardtii* (XP\_001690111.1); Csc, *Centruroides sculpturatus* (XP\_023219872.1); Csu, *Coccomyxa subellipsoidea* (XP\_005645781.1); Cme1, *Cyanidioschyzon merolae* (XP\_005538077.1); Cme2, *Cyanidioschyzon merolae* (XP\_005536888.1); Cva1, *Chlorella variabilis* (XP\_005848966.1); Cva2, *Chlorella variabilis* (XP\_005847218.1); Cva3, *Chlorella variabilis* (XP\_005851997.1); Cvi, *Crassostrea virginica* (XP\_022340710.1); Lpo, *Limulus polyphemus* (XP\_013787955.2); Mpu, *Micromonas pusilla* (XP\_003057128.1); Mye, *Mizuhopecten yessoensis* (XP\_021349948.1); Nme, *Numida meleagris* (XP\_021244254.1); SaCLCc1, *Suaeda altissima* (AVQ93350.1); Vca1, *Volvox carteri* (XP\_002953983.1); Vca2, *Volvox carteri* (XP\_002953895.1). Белки *Arabidopsis thaliana*: AtCLCa (NP\_198905.1), AtCLCb (NP\_189353.1), AtCLCc (NP\_199800.1), AtCLCd (NP\_197996.1), AtCLCe (NP\_001190924.1), AtCLCf (NP\_564698.1), AtCLCg (NP\_198313.2).

Для поиска в файлах с контигами последовательностей, кодирующих белки семейства CLC, в качестве запроса использовали протяженные вырожденные

консенсусные последовательности, полученные при выравнивании последовательностей, кодирующих белки CLC различных организмов, прежде всего зеленых водорослей. Эти последовательности были выбраны биоинформатическим поиском в базах данных возможных последовательностей, кодирующих белки CLC у водорослей.

В собранных транскриптомах были *in silico* идентифицированы контиги, содержащие открытые рамки считывания для трех различных белков, гомологичных белкам семейства CLC, которые мы обозначили как DtCLC1 (857 а.о.), DtCLC2 (809 а.о.), DtCLC3 (768 а.о.).

На рисунке представлено филогенетическое дерево принадлежащих семейству CLC белков из различных организмов, включая идентифицированные *in silico* белки CLC водоросли *D. tertiolecta*. Прежде всего, следует отметить, что найденные белки *D. tertiolecta* высокогомологичны белкам CLC из других водорослей. Один из идентифицированных белков, DtCLC1, сходен с белками высших растений, в частности, с белком AtCLCd из *A.thaliana*, который, как предполагается, является Cl<sup>-</sup>/H<sup>+</sup> антипортером и функционирует в транс-Гольджи сети [Lv et al., 2009]. Два других белка, DtCLC2 и DtCLC3, лежат в одной кладе с белками CLC животных (прежде всего рыб и рептилий), также осуществляющими Cl<sup>-</sup>/H<sup>+</sup> обмен. Известно, что белки CLC, перенося Cl<sup>-</sup> через клеточные мембраны, контролируют электрический потенциал на мембранах, pH и состав электролитов во внутриклеточных органеллах [Jentsch, 2008]. Результаты нашего теоретического поиска представляют основу для дальнейших экспериментальных исследований белков CLC у водорослей, выяснения их клеточной локализации и физиологической роли у этих организмов.

#### Литература

Балнокин Ю.В., Медведев А.В., Калашникова Т.С., Галкина И.В. Ионный гомеостаз в цитозоле одноклеточных водорослей при засолении среды хлористым натрием // Журнал общей биологии. – 1990 – Т. 51, № 2. – С. 234–246.

Масюк Н.П. Морфология, систематика экология, географическое распространение рода *Dunaliella* и перспективы его практического использования. – Киев, Наукова думка, 1973. – 243–244 с.

Попова Л.Г., Балнокин Ю.В. Na<sup>+</sup>-АТФазы галотолерантных микроводорослей // Физиология растений. – 2013 – Т. 60, № 4. – С. 499–511.

Altschul S.F., Madden T.L., Schäffer A.A., Zhang J., Zhang Z., Miller W., Lipman D.J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs // Nucl. Acids Res. – 1997. – V. 25. – P. 3389–3402.

Bolger A.M., Lohse M., Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data // Bioinformatics. – 2014. – btu170.

Feng L., Campbell E.B., Hsiung Y., MacKinnon R. Structure of a eukaryotic CLC transporter defines an intermediate state in the transport cycle // Science. – 2010 – V. 330. – P. 635–641.

Haas B.J., Papanicolaou A., Yassour M., Grabherr M., Blood P.D. et al. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis // Nat. protocols. – 2013 – V. 8. – P. 1494–1512.

Jentsch T.J. CLC chloride channels and transporters: from genes to protein structure, pathology and physiology // Critical reviews in biochemistry and molecular biology. – 2008 – V. 43. – P. 3–36.

Kirst G.O. Ion composition of unicellular marine and fresh-water algae, with special reference to *Platymonas subcordiformis* cultivated in media with different osmotic strengths // Oecologia (Berl.). – 1977. – V. 28. – P. 177–189.

Langmead B., Salzberg S.L. Fast gapped-read alignment with Bowtie 2 // Nat. Methods. – 2012. – V. 9. – P. 357–359.

Lv Q.D., Tang R.J., Liu H., Gao X.S., Li Y.Z., Zheng H.Q., Zhang H.X. Cloning and molecular analyses of the *Arabidopsis thaliana* chloride channel gene family // Plant Science. – 2009 – V. 176. – P. 650–661.

Munns R., Tester M. Mechanisms of salinity tolerance // Annu. Rev. Plant Biol. – 2008. – V. 59. – P. 651–681.

Teakle N.L., Tyerman S.D. Mechanisms of Cl<sup>−</sup> transport contributing to salt tolerance // Plant, cell, environment. – 2010 – V. 33. – P. 566–589.

### ***IN SILICO* IDENTIFICATION OF SEQUENCES ENCODING CLC FAMILY PROTEINS IN THE *DE NOVO* ASSEMBLED TRANSCRIPTOMES OF MARINE ALGA *DUNALIELLA TERTIOLECTA***

L.G. Popova<sup>1</sup>, A.V. Shuvalov<sup>1</sup>, A.A. Yurchenko<sup>2</sup>, D.E. Khramov<sup>1,3</sup>, D.A. Matalin<sup>1</sup>, L.A. Khalilova<sup>1</sup>, Y.V. Orlova<sup>1</sup>, Y.V. Balnokin<sup>1,3</sup>

<sup>1</sup>K.A. Timiryazev Institute of Plant Physiology RAS, Moscow, Russia, [lora\\_gp@mail.ru](mailto:lora_gp@mail.ru)

<sup>2</sup>Federal State Budgetary Educational Institution of Higher Education "Saint-Petersburg State University", Theodosius Dobzhansky Center for Genome Bioinformatics, Saint-Petersburg, Russia

<sup>3</sup>Federal State Budget Educational Institution of Higher Education M.V. Lomonosov Moscow State University, Moscow, Russia

**Abstract.** For microalgae *Dunaliella tertiolecta* *de novo* transcriptome assemblies were accomplished using the short-read RNA-seq data deposited in the Sequence Read Archive database (NCBI). In the assembled transcriptomes by *in silico* analysis we found sequences encoding the Cl<sup>−</sup>-transporting proteins of the CLC family: *DtCLC1* (857 a.a.), *DtCLC2* (809 a.a.), *DtCLC3* (768 a.a.). These proteins are similar to the CLC proteins of higher plants and animals (primarily fish and reptiles) that fulfil Cl<sup>−</sup>/H<sup>+</sup> exchange.

**Keywords:** *de novo* transcriptome assembly, CLC proteins, marine microalgae, *Dunaliella tertiolecta*