

БІОІНФОРМАТИЧНИЙ АНАЛІЗ КОДУВАЛЬНОЇ ДІЛЯНКИ sHSP У *NICOTIANA SYLVESTRIS*

Л.В. КОЗУБ, І.І. ПАНЧУК

Кафедра молекулярної генетики та біотехнології,
Чернівецький національний університет імені Юрія Федьковича
вул. Коцюбинського, 2, м. Чернівці, Україна, 58012,
e-mail: i.panchuk@chnu.edu.ua

Протягом еволюції у рослин виробились механізми захисту від температурного стресу, які пов'язані із синтезом захисних білків, серед яких переважають молекулярні шаперони, зокрема – низькомолекулярні білки теплового стресу (sHSP). Незважаючи на важливу роль, яку ці білки відіграють у захисній реакції на високотемпературний стрес, вони залишаються все ще недостатньо дослідженими для багатьох таксономічних груп рослин. Зокрема, про організацію та різноманіття генів/білків теплового шоку у представників такого важливого роду як *Nicotiana* (тютюн) майже нічого невідомо. У статті наводяться результати біоінформатичного аналізу мультигенної / мультипротеїнової родини sHSP у *N. sylvestris*, важливого модельного об'єкту у фізіології, біохімії, молекулярній генетиці та клітинній біотехнології рослин.

Для пошуку гомологічних послідовностей використовували базу даних GenBank із застосуванням алгоритму BLAST. Амінокислотні послідовності вирівнювали методом L-INS-I на сервері MAFFT. Філогенетичний аналіз проводили методом Maximum likelihood за допомогою плагіна PhyML для Geneious Prime 2023.2.1. Статистична підтримка гілок була розрахована за методом aLRT Chi2.

Встановлено, що в межах порівнюваних амінокислотних послідовностей sHSP *N. sylvestris* можна виділити три частини: мінливу N-кінцеву область, консервативний α -кристаліновий домен (ACD) та C-кінцеву ділянку. На основі вирівнювання амінокислотних послідовностей sHSP було побудовано дендрограму подібності, на якій з високою статистичною підтримкою візуалізуються декілька клад.

Результати порівняльного аналізу амінокислотних послідовностей свідчать, що sHSP *N. sylvestris* належать до 10 структурних класів. Білки семи класів імовірно локалізуються у цитоплазмі та/або в клітинному ядрі, тоді як решта – у ендоплазматичному ретикулумі, мітохондріях, пластидях та пероксисомах.

На загал за результатами біоінформатичного аналізу у геномі *N. sylvestris* виявлено 24 гени, які кодують sHSP та один псевдоген, у якого втрачено фрагмент, що кодує N-кінцеву область. Враховуючи, що послідовність псевдогена незначно відрізняється від найближчих представників цієї групи, можна припустити, що перетворення цієї ділянки у псевдоген є порівняно нещодавньою еволюційною подією.

Ключові слова: абіотичний стрес, біоінформатичний аналіз, генетичний поліморфізм, молекулярна геноміка, молекулярна еволюція, мультигенна родина, α -кристаліновий домен, sHSP, *N. sylvestris*.

Вступ. Рослини ведуть прикріплений спосіб життя у природньому середовищі, де зазнають різноманітних абіотичних та біотичних стресів. Підвищення температури, пов'язане з глобальним потеплінням, негативно впливає на ріст і розвиток рослин. Щоб простояти шкідливій дії високої температури та вижити, у рослин протягом еволюції виникли механізми захисту від високотемпературного стресу, які пов'язані із експресією чисельних стресових генів та синтезом захисних білків, серед яких переважають молекулярні шаперони (Waters & Vierling, 2020) та білки антиоксидантного захисту (Panchuk et al., 2002; Пиріжок та ін., 2009; Руснак та ін., 2013; Буздуга та ін., 2014, Poudel & Poudel, 2020). У рослин білки-шаперони представлені кількома мультипротеїновими родинами білків теплового шоку (heat shock proteins, HSP), які різняться молекулярною масою (Khan et al., 2021).

Важлива роль належить малим (низькомолекулярним) HSP (small HSP, sHSP), які є АТФ-незалежними молекулярними шаперонами і запобігають необоротній агрегації денатурованих за дії високої температури білків (Waters, 2012; Waters & Vierling, 2020). sHSP – це велика група білків, що мають висококонсервативний α -кристаліновий домен (ACD), коротку C-кінцеву послідовність та відносно велику варіабельну N-кінцеву ділянку. N-кінець бере участь у зв'язуванні з денатурованими білками, тоді як C-кінець залучений до гомоолігомеризації через консервативний мотив та утворення гранул теплового стресу (Hagymasi et al., 2022; Hassan et al., 2020). На сьогоднішній день ідентифіковано одинадцять (Waters & Vierling, 2020) мультипротеїнових родин sHSP як у однодольних, так і у дводольних рослин. Представники п'яти з

цих родин локалізовані в цитоплазмі, а інших шести – у клітинних органелах, включаючи ядро, хлоропласти, мітохондрії, ендоплазматичний ретикулум та пероксисоми. Різна внутрішньоклітинна локалізація може вказувати на функціональну дивергенцію між цими сімействами sHSP (Waters & Vierling, 2020; Zhang et al., 2015).

Nicotiana sylvestris являє собою диплоїдний вид ($2n = 2x = 24$), який походить з регіону Анд в Аргентині та Болівії у Південній Америці. *N. sylvestris* культивується здебільшого як декоративна рослина, а також використовується як модельний організм у дослідженнях синтезу терпеноїдів (Ennajdaoui et al., 2010; Sallaud et al., 2012), інженерії пластид (Maliga & Svab, 2010; Thyssen et al., 2012), стійкості до біотичного та абіотичного стресу (Sekine et al., 2012; Sierro et al., 2013).

N. sylvestris є також популярним об'єктом у дослідженнях поліплоїдного видоутворення, оскільки він є донором материнського субгеному алотетраплоїдного ($2n = 4x = 48$) *N. tabacum*, який виник внаслідок гібридизації між *N. sylvestris* та *N. tomentosiformis* ($2n = 2x = 24$) (Okamuro & Goldberg, 1985; Goodspeed & Thompson, 1959; Miroshnichenko et al., 1988; D'Andrea et al., 2023).

На сьогодні для представників родини Пасльонові (Solanaceae) механізми захисної відповіді на високотемпературний стрес добре досліджені переважно для томату, *Solanum lycopersicum* та його диких родичів (Scharf et al., 1998; Alsamir et al., 2021; Aldubai et al., 2022), тоді як про організацію та різноманіття генів/білків теплового шоку у представників такого важливого роду як *Nicotiana* майже нічого невідомо (Volkov et al., 2005). У цій статті ми наводимо результати біоінформатичного аналізу мультигенної / мультипротеїнової родини sHSP у *N. sylvestris*.

Матеріали та методи. Для пошуку гомологічних послідовностей ми використовували базу даних GenBank із застосуванням алгоритму BLAST (Boratyn et al., 2013).

Амінокислотні послідовності вирівнювали методом L-INS-I на сервері MAFFT (Katoh et al., 2019). Філогенетичний аналіз проводили методом Maximum likelihood за допомогою плагіна PhyML для Geneious Prime 2023.2.1 (Guindon & Gascuel, 2003). Статистична підтримка гілок була розрахована за методом aLRT Chi2 (Anisimova & Gascuel, 2006). Розраховане дерево експортували у форматі Newick та візуалізували з використанням онлайн інструменту iTOL v6 - Interactive tree of life (Letunic & Bork, 2021).

Результати та їх обговорення. Біоінформатичний скринінг дозволив нам ідентифікувати у геномі *N. sylvestris* 25 послідовностей, які мають високу подібність із кодувальною ділянкою (coding sequence - CDS) генів sHSP. Ці ділянки анотовані у GenBank як передбачені (predicted) гени sHSP. Загальна довжина амінокислотної послідовності цих білків (за виключенням NsHSP12.0-CI – див. нижче) знаходиться у межах від 137 (NsHSP15.7-CV) до 241 (NsHSP26.4-P) АК, а розрахована молекулярна маса – від 15.7 до 26.4 kDa, відповідно (табл. 1).

Всі ідентифіковані нами у геномі *N. sylvestris* послідовності були порівняні між собою, а також із описаними раніше sHSP двох видів томату, *Solanum lycopersicum* L. (*Lycopersicon esculentum* (L.) Mill.) та *S. peruvianum* L. (*L. peruvianum* (L.) Mill.), які належить до тієї ж самої родини Solanaceae (рис. 1). Встановлено, що в межах порівнюваних амінокислотних послідовностей можна виділити три частини: мінливу N-кінцеву область, консервативний α -кристаліновий домен (ACD) та C-кінцеву ділянку.

На основі вирівнювання амінокислотних послідовностей sHSP було побудовано дендрограму подібності, на якій з високою статистичною підтримкою візуалізуються декілька клад (рис. 2).

Наступним кроком аналізу було визначення структурного класу, до якого належить той чи інший білок. При цьому ми орієнтувалися на пептидні сигнали клітинного транспорту/локалізації, присутні у N-кінцевій області та C-кінцевій ділянці та враховували анотацію послідовностей, які наводяться у базі даних GenBank. Крім того, ми брали до уваги позицію кожного білка на дендрограмі, в першу чергу – порівняно із розташуванням sHSP томату, для яких відома приналежність до певного структурного класу.

Результати аналізу свідчать, що sHSP *N. sylvestris* належать до 10 структурних класів (табл. 1). Білки семи класів імовірно локалізуються у цитоплазмі та/або в клітинному ядрі, тоді як решта – у мітохондріях, пластидах та пероксисомах. Слід зазначити, що визначена приналежність досліджуваних білків до певного класу ґрунтується виключно на результатах біоінформатичного аналізу і потребує експериментальної перевірки.

Довжина N-кінцевої області цитоплазматичних sHSP варіює від 30 до 90 АК. Найдовша амінокислотна послідовність спостерігається у класу CI і складає 90 АК, тоді як у класу CV вона найкоротша – 31 АК. У N-кінцевій області пластидних sHSP знаходиться від 120 до 145 АК,

мітохондріальних sHSP – від 104 до 130 залишків, а пероксисомальних – лише 28 амінокислотних залишків. Альфа-кристаліновий домен містить близько 90 амінокислотних залишків. С-кінцева ділянка є короткою: вона містить від 14 до 33 АК у цитоплазматичних, 13 у пластидних, 14-16 у

мітохондріальних і 18 у пероксисомального sHSP (рис. 1). До класу CI (cytoplasmic I) належать вісім sHSP *N. sylvestris*, з яких сім є дуже подібними між собою та із цитоплазматичним білком томату LpHSP17.7-CI (CAA12388), тоді як білок NsHSP22.2-CI дещо відрізняється від всіх інших.

Таблиця 1.
Характеристика нуклеотидних та амінокислотних послідовностей sHSP *Nicotiana sylvestris*
Table 1.

Characteristics of nucleotide and amino acid sequences of *Nicotiana sylvestris* sHSP

№	Назва гена/білка	Довжина CDS		Молекулярна маса, kDa	Номер нуклеотидної послідовності	Номер амінокислотної послідовності	Анотація гену у GenBank
		нп	АК				
Клас CI							
1	NsHSP17.9-CI	480	159	17.9	XM_009773397.1	XP_009771699.1	<i>N. sylvestris</i> 17.6 kDa class I HSP 3-like
2	NsHSP17.5-CI	462	153	17.5	XM_009764599.1	XP_009762901.1	<i>N. sylvestris</i> 17.6 kDa class I HSP-like
3	NsHSP17.5-CI	462	153	17.5	XM_009763174.1	XP_009761476.1	<i>N. sylvestris</i> 17.6 kDa class I HSP-like
4	NsHSP19.9-CI	534	177	19.9	XM_009795322.1	XP_009793624.1	<i>N. sylvestris</i> 17.8 kDa class I HSP -like
5	NsHSP18.1-CI	480	159	18.1	XM_009781855.1	XP_009780157.1	<i>N. sylvestris</i> 17.8 kDa class I HSP -like
6	NsHSP17.5-CI	465	154	17.5	XM_009772066.1	XP_009770368.1	<i>N. sylvestris</i> 17.8 kDa class I HSP-like
7	NsHSP22.2-CI	588	195	22.2	XM_009769562.1	XP_009767864.1	<i>N. sylvestris</i> 18.1 kDa class I HSP -like
8	NsHSP12.0-CI	315	104	12.0	XM_009774444.1	XP_009772746.1	<i>N. sylvestris</i> 18.1 kDa class I HSP-like
Клас CII							
9	NsHSP21.5-CII	588	195	21.5	XM_009767172.1	XP_009765474.1	<i>N. sylvestris</i> 17.1 kDa class II HSP-like
10	NsHSP17.5-CII	474	157	17.5	XM_009796403.1	XP_009794705.1	<i>N. sylvestris</i> 17.3 kDa class II HSP-like
11	NsHSP17.4-CII	474	157	17.4	XM_009760536.1	XP_009758838.1	<i>N. sylvestris</i> 17.3 kDa class II HSP-like
12	NsHSP18.3-CII	486	161	18.3	XM_009767173.1	XP_009765475.1	<i>N. sylvestris</i> 18.8 kDa class II HSP-like
Клас CIII							
13	NsHSP16.7-CIII	453	150	16.7	XM_009802241.1	XP_009800543.1	<i>N. sylvestris</i> 17.4 kDa class III HSP

Клас CIV							
14	NsHSP16.7-CIV	579	192	21.7	XM_009765184.1	XP_009763486.1	<i>N. sylvestris</i> 22.0 kDa class IV HSP-like
15	NsHSP21.6-CIV	573	190	21.6	XM_009760546.1	XP_009758848.1	<i>N. sylvestris</i> 16.9 kDa class I HSP 1-like
Клас CV							
16	NsHSP15.7-CV	414	137	15.7	XM_009764074.1	XP_009762376.1	<i>N. sylvestris</i> 15.4 kDa class V HSP
Клас CVI							
17	NsHSP17.7-CVI	480	159	17.7	XM_009781349.1	XP_009779651.1	<i>N. sylvestris</i> 17.1 kDa class II HSP-like
Клас CVII							
18	NsHSP18.4-CVII	495	164	18.4	XM_009772008.1	XP_009770310.1	<i>N. sylvestris</i> 16.9 kDa class I HSP 1-like
19	NsHSP18.5-CVII	495	164	18.5	XM_009797062.1	XP_009795364.1	<i>N. sylvestris</i> 16.9 kDa class I HSP 1-like
Клас P							
20	NsHSP26.4-P	726	241	26.4	XM_009793483.1	XP_009791785.1	<i>N. sylvestris</i> sHSP, chlorplastic
21	NsHSP24.9-P	654	217	24.9	XM_009781309.1	XP_009779611	<i>N. sylvestris</i> HSP, chlorplastic-like
Клас M							
22	NsHSP21.3-M	579	192	21.3	XM_009764739.1	XP_009763041.1	<i>N. sylvestris</i> 22 kDa HSP, mitochondrial-like
23	NsHSP25.7-M	693	230	25.7	XM_009788708.1	XP_009787010.1	<i>N. sylvestris</i> 26.5 kDa HSP, mitochondrial-like
24	NsHSP24.4-M	651	216	24.4	XM_009764740.1	XP_009763042	<i>N. sylvestris</i> sHSP, chlorplastic-like
Клас Per							
25	NsHSP15.8-Per	432	143	15.8	XM_009781403.1	XP_009779705.1	<i>N. sylvestris</i> 15.7 kDa HSP, peroxisomal

Варто також зазначити, що до класу CI також належить білок NsHSP12.0-CI. Проте, аналіз цієї послідовності показує, що вона відрізняється від решти делецією N-кінцевої області (включно із стартовим метіоніновим кодоном) та перших трьох амінокислот ACD. Відповідно, можна висунути припущення, що ця послідовність швидше за все являє собою нефункціонуючий псевдоген. Враховуючи, що послідовність NsHSP12.0-CI лише незначно відрізняється від

найближчих представників цієї групи, можна припустити, що делеція N-кінцевої області та перетворення цієї ділянки у псевдоген відбулося порівняно нещодавно в еволюції.

До класу CII належать чотири білки, які є високоподібними між собою. До цієї групи також входить білок томату LpHSP17.3-CII (CAA12390).



Рис. 1. Порівняння амінокислотних послідовностей sHSP для *Nicotiana sylvestris* (Ns), *Solanum lycopersicum* (*Lycopersicon esculentum*, Le) and *S. peruvianum* (*L. peruvianum*, Lp). Стрілками вказано позиції потенційних вторинних структурних елементів у межах АСД.

Fig. 1. Amino acid sequence comparison of sHSP of *Nicotiana sylvestris* (Ns), *Solanum lycopersicum* (*Lycopersicon esculentum*, Le) and *S. peruvianum* (*L. peruvianum*, Lp).

Arrows indicate the positions of potential secondary structural elements within the ACD.

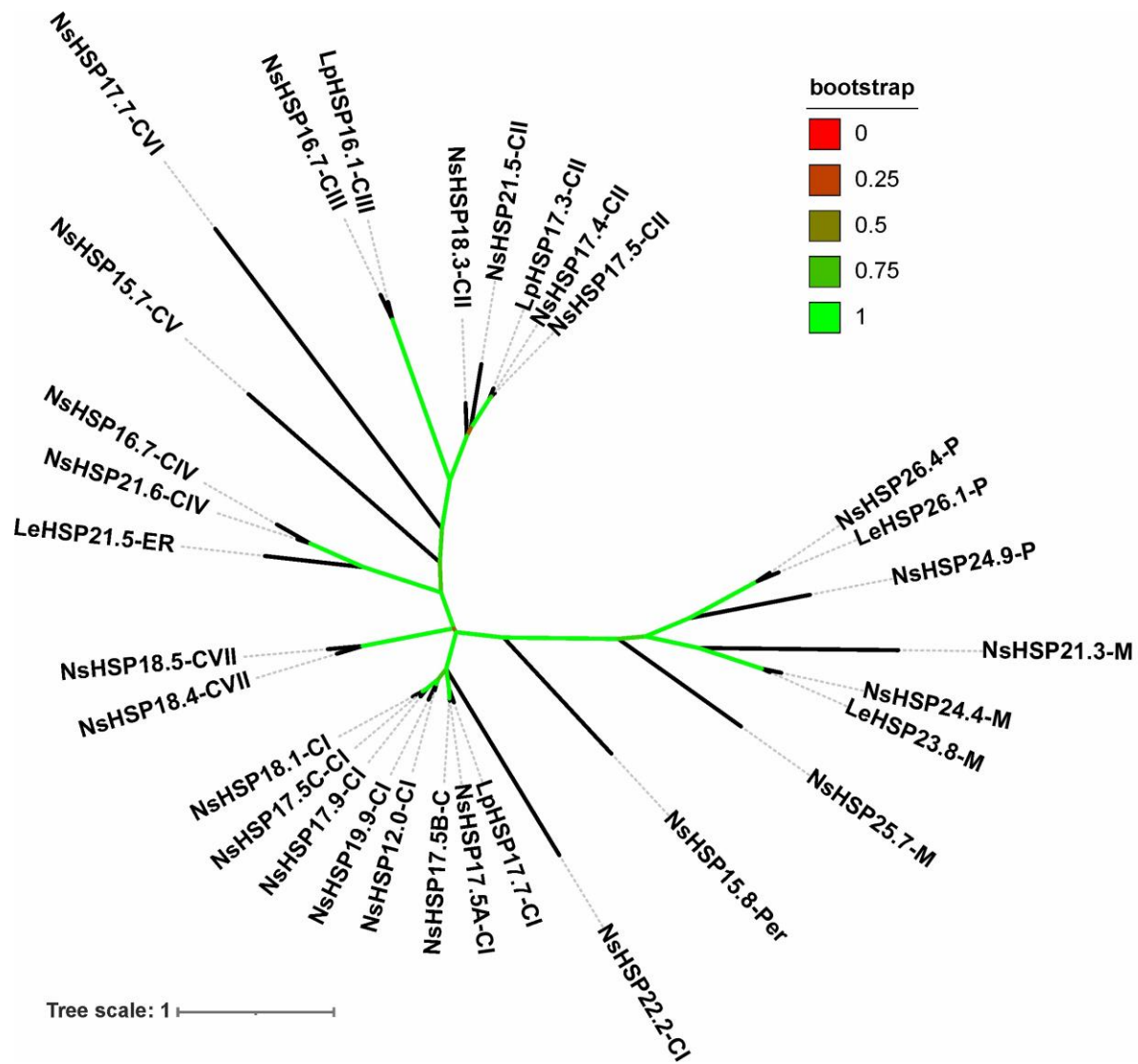


Рис. 2. ML-дендрограма подібності амінокислотних послідовностей sHSP *Nicotiana sylvestris*.
 Fig. 2. ML-dendrogram reflecting amino acid sequence similarity of *Nicotiana sylvestris* sHSP.

До класів CIII, CIV та CVI належать лише по одному білку, з яких білок NsHSP16.7-CIII є високоподібним до LpHSP16.1-CIII (AAK84869) томату. Хоча білки цих трьох класів помітно відрізняються між собою та від інших цитоплазматичних sHSP, вони формують спільну гілку на дендрограмі, що свідчить про їх дивергенцію від спільного предкового білка.

До класу CIV у *N. sylvestris* належить два білки, які формують на дендрограмі спільну гілку із LeHSP21.5-ER (BAA97658) томата, який транспортується в ендоплазматичний ретикулум. Відповідно, можна припустити, що білки класу CIV у *N. sylvestris* також можуть локалізуватись в ендоплазматичному ретикулумі. Проте, детальне порівняння послідовностей N- та С-кінцевих ділянок, які відповідають за внутрішньоклітинний транспорт білкових молекул, свідчить про наявність суттєвих відмінностей між NsHSP16.7-CIV, NsHSP21.6-CIV та LeHSP21.5-ER. Отже, клітинна локалізація sHSP класу CIV у *N. sylvestris* потребує подальшого уточнення.

Два високоподібні sHSP класу CVII *N. sylvestris* утворюють окрему гілку на дендрограмі. Гомологічних білків у томату не знайдено.

У геномі *N. sylvestris* було також ідентифіковано ділянки, які кодують три мітохондріальні та два пластидні sHSP. Один мітохондріальний та один пластидний білки, NsHSP24.4-M та NsHSP26.4-P, виявилися високоподібними до білків томату, LeHSP23.8-M (NP_001233872) та LeHSP26.1-P (AAB49626), відповідно.

У нашому біоінформатичному скринінгу було також ідентифіковано білок NsHSP15.8-Per, який містить на С-кінці відомий сигнал локалізації у пероксисомах, SKL. Цікаво, що для томату пероксисомальні sHSP не описані.

Послідовності мітохондріальних, хлоропластних та пероксисомальних sHSP утворюють спільну гілку на дендрограмі.

Висновки. За результатами біоінформатичного аналізу у геномі *N. sylvestris* присутні 24 гени, які кодують sHSP та один псевдоген. Ідентифіковані білки мають типову для sHSP структуру: вони складаються з мінливої N-кінцевої ділянки, консервативного α -кристалінового домену та С-кінцевої ділянки. Порівняння амінокислотних послідовностей показало, що sHSP *N. sylvestris* мають локалізуватись у цитоплазмі/ядрі, ендоплазматичному ретикулумі, мітохондріях, пластидах та пероксисомах. Цитоплазматичні sHSP є найбільш чисельними.

Подяки. Автори висловлюють щирі подяки професору Р.А. Волкову за участь в обговоренні отриманих результатів та ст. наук. співр.

Ю.О. Тинкевичу за корисні поради при проведенні біоінформатичного скринінгу.

Список літератури / References:

1. Aldubai, A. A., Alsadon, A. A., Migdadi, H. H., Alghamdi, S. S., Al-Faifi, S. A., & Afzal, M. (2022). Response of tomato (*Solanum lycopersicum* L.) genotypes to heat stress using morphological and expression study. *Plants*, 11(5), 615. <https://doi.org/10.3390/plants11050615>
2. Alsamir, M., Mahmood, T., Trethowan, R., & Ahmad, N. (2021). An overview of heat stress in tomato (*Solanum lycopersicum* L.). *Saudi Journal of Biological Sciences*, 28(3), 1654–1663. <https://doi.org/10.1016/j.sjbs.2020.11.088>
3. Anisimova, M., & Gascuel, O. (2006). Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Systematic Biology*, 55(4), 539–552. <https://doi.org/10.1080/1063515060075545>
4. Boratyn, G. M., Camacho, C., Cooper, P. S., Coulouris, G., Fong, A., Ma, N., Madden, T. L., Matten, W. T., McGinnis, S. D., Merezuk, Y., Raytselis, Y., Sayers, E. W., Tao, T., Ye, J., & Zaretskaya, I. (2013). BLAST: a more efficient report with usability improvements. *Nucleic Acids Research*, 41(W1), W29–W33. [tps://doi.org/10.1093/nar/gkt282](https://doi.org/10.1093/nar/gkt282)
5. Buzduga, I.M., Volkov, R.A., & Panchuk I.I. (2014). Heat stress affects lipid peroxidation and activity of ascorbate peroxidase in *Nicotiana tabacum* [Vplyv teplovoho stresu na peroksidne okyslennia lipidiv ta activnist askorbat peroksidasy u *Nicotiana tabacum*]. *Physiologia Rastenij i Genetika*. 46 (2), 151-157. [In Ukrainian].
6. D'Andrea, L., Sierro, N., Ouadi, S., Hasing, T., Rinaldi, E., Ivanov, N. V., & Bombarely, A. (2023). Polyploid *Nicotiana* section *Suaveolentes* originated by hybridization of two ancestral *Nicotiana* clades. *Frontiers in Plant Science*, 14. <https://doi.org/10.3389/fpls.2023.999887>
7. Ennajdaoui, H., Vachon, G., Giacalone, C., Besse, I., Sallaud, C., Herzog, M., & Tissier, A. (2010). Trichome specific expression of the tobacco (*Nicotiana sylvestris*) cembratrien-ol synthase genes is controlled by both activating and repressing cis-regions. *Plant Molecular Biology*, 73(6), 673–685. <https://doi.org/10.1007/s11103-010-9648-x>
8. González-Gordo, S., Palma, J., & Corpas, F. (2023). Small heat shock protein (sHSP) gene family from sweet pepper (*Capsicum annuum* L.) fruits: involvement in ripening and modulation by nitric oxide (NO). *Plants*, 12(2), 389. <https://doi.org/10.3390/plants12020389>
9. Goodspeed, T. H., & Mildred C. Thompson. (1959). Cytotaxonomy of *Nicotiana*. II. *Botanical Review*, 25(2), 385–415. <http://www.jstor.org/stable/4353598>
10. Guindon, S., & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, 52(5), 696–704. <https://doi.org/10.1080/10635150390235520>
11. Hagymasi, A. T., Dempsey, J. P., & Srivastava, P. K. (2022). Heat-Shock proteins. *Current Protocols*, 2(11). <https://doi.org/10.1002/cpz1.592>

12. Hassan, M. U., Chattha, M. U., Khan, I., Chattha, M. B., Barbanti, L., Aamer, M., Iqbal, M. M., Nawaz, M., Mahmood, A., Ali, A., & Aslam, M. T. (2020). Heat stress in cultivated plants: nature, impact, mechanisms, and mitigation strategies—a review. *Plant Biosystems*, 155(2), 211–234. <https://doi.org/10.1080/11263504.2020.1727987>
13. Ji, X., Yu, Y., Ni, P., Zhang, G., & Guo, D. (2019). Genome-wide identification of small heat-shock protein (HSP20) gene family in grape and expression profile during berry development. *BMC Plant Biology*, 19(1). <https://doi.org/10.1186/s12870-019-2031-4>
14. Katoh, K., Rozewicki, J., & Yamada, K. D. (2017). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, 20(4), 1160–1166. <https://doi.org/10.1093/bib/bbx108>
15. Khan, S., Jabeen, R., Deeba, F., Waheed, U., Khanum, P., & Iqbal, N. (2021). Heat shock proteins: classification, functions and expressions in plants during environmental stresses. *Journal of Bioresource Management*, 8(2), 9. <https://doi.org/10.35691/JBM.1202.0183>
16. Letunic, I., & Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, 49(W1), W293–W296. <https://doi.org/10.1093/nar/gkab301>
17. Maliga, P., & Svab, Z. (2010). Engineering the plastid genome of *Nicotiana glauca*, a diploid model species for plastid genetics. In *Methods in molecular biology* (pp. 37–50). https://doi.org/10.1007/978-1-61737-957-4_2
18. Miroshnichenko, G. P., Volkov, R. A., & Kostishin, S. S. (1988). Polynucleotide sequence divergence in DNAs of interspecific Solanaceae hybrids. *Biochemistry (Moscow)*, 53, 565–572.
19. Okamoto, J. K., & Goldberg, R. B. (1985). Tobacco single-copy DNA is highly homologous to sequences present in the genomes of its diploid progenitors. *Molecular and General Genetics MGG*, 198(2), 290–298. <https://doi.org/10.1007/BF00383009>
20. Panchuk, I. I., Volkov, R. A., & Schöffl, F. (2002). Heat stress-and heat shock transcription factor-dependent expression and activity of ascorbate peroxidase in Arabidopsis. *Plant physiology*, 129(2), 838–853. <https://doi.org/10.1104/pp.001362>
21. Poudel, P. B., & Poudel, M. R. (2020). Heat stress effects and tolerance in wheat: A review. *J. Biol. Today's World*, 9(3), 1–6.
22. Pyrizhok, R. Yu., Volkov, R. A., & Panchuk I. I. (2009). Peroxidase activity in maize seedlings upon heat stress [Активність пероксидази проростків кукурудзи в умовах теплового стресу] *Physiologia i biochimia kulturnykh rastenij*. 41 (1): 44–49. [In Ukrainian].
23. Rusnak, T. O., Doliba, I. M., Volkov, R. A., & Panchuk I. I. (2013). Guaiacol peroxidase activity in Cat2 knock-out mutant of *Arabidopsis thaliana* upon heat stress treatment. [Активність гуаїаколу пероксидази у Cat2 knock-out mutantiv *Arabidopsis thaliana* за дії теплового стресу]. *Physiologia i biochimia kulturnykh rastenij*. 45 (3), 246–253. [In Ukrainian].
24. Sallaud, C., Giacalone, C., Töpfer, R., Goepfert, S., Bakaher, N., Rösti, S., & Tissier, A. (2012). Characterization of two genes for the biosynthesis of the labdane diterpene Z-abienol in tobacco (*Nicotiana tabacum*) glandular trichomes. *Plant Journal*, 72(1), 1–17. <https://doi.org/10.1111/j.1365-313x.2012.05068.x>
25. Scharf, K. D., Höfheld, I., & Nover, L. (1998). Heat stress response and heat stress transcription factors. *Journal of biosciences*, 23, 313–329.
26. Scharf, K., Siddique, M., & Vierling, E. (2001). The expanding family of *Arabidopsis thaliana* small heat stress proteins and a new family of proteins containing α -crystallin domains (Acid proteins). *Cell Stress & Chaperones*, 6(3), 225. [https://doi.org/10.1379/1466-1268\(2001\)006](https://doi.org/10.1379/1466-1268(2001)006)
27. Sekine, K., Tomita, R., Takeuchi, S., Atsumi, G., Saitoh, H., Mizumoto, H., Kiba, A., Yamaoka, N., Nishiguchi, M., Hikichi, Y., & Kobayashi, K. (2012). Functional differentiation in the Leucine-Rich repeat domains of closely related plant Virus-Resistance proteins that recognize common AVR proteins. *Molecular Plant-microbe Interactions*, 25(9), 1219–1229. <https://doi.org/10.1094/mpmi-11-11-0289>
28. Sierro, N., Battey, J. N., Ouali, S., Bovet, L., Goepfert, S., Bakaher, N., Peitsch, M. C., & Ivanov, N. V. (2013). Reference genomes and transcriptomes of *Nicotiana glauca* and *Nicotiana glauca*. *Genome Biology*, 14(6). <https://doi.org/10.1186/gb-2013-14-6-r60>
29. Thyssen, G., Svab, Z., & Maliga, P. (2012). Exceptional inheritance of plastids via pollen in *Nicotiana glauca* with no detectable paternal mitochondrial DNA in the progeny. *Plant Journal*, 72(1), 84–88. <https://doi.org/10.1111/j.1365-313x.2012.05057.x>
30. Volkov, R. A., Panchuk, I. I., & Schöffl, F. (2005). Small heat shock proteins are differentially regulated during pollen development and following heat stress in tobacco. *Plant Molecular Biology*, 57, 487–502. <https://doi.org/10.1007/s11103-005-0339-y>
31. Waters, E. R. (2012). The evolution, function, structure, and expression of the plant sHSPs. *Journal of Experimental Botany*, 64(2), 391–403. <https://doi.org/10.1093/jxb/ers355>
32. Waters, E. R., & Vierling, E. (2020). Plant small heat shock proteins – evolutionary and functional diversity. *New Phytologist*, 227(1), 24–37. <https://doi.org/10.1111/nph.16536>
33. Yemets, A., Stelmakh, O., & Blume, Y. B. (2008). Effects of the herbicide isopropyl-N-phenyl carbamate on microtubules and MTOCs in lines of *Nicotiana glauca* resistant and sensitive to its action. *Cell Biology International*, 32(6), 623–629. <https://doi.org/10.1016/j.cellbi.2008.01.012>
34. Zhang, J., Chen, H., Wang, H., Li, B., Yi, Y., Kong, F., Liu, J., & Zhang, H. (2015). Constitutive expression of a tomato small heat shock protein gene LEHSP21 improves tolerance to high-temperature stress by enhancing antioxidation capacity in tobacco. *Plant Molecular Biology Reporter*, 34(2), 399–409. <https://doi.org/10.1007/s11105-015-0925-3>

BIOINFORMATIC ANALYSIS OF THE CODING SEQUENCES OF *NICOTIANA SYLVESTRIS* sHSP

L. V. Kozub, I. I. Panchuk

*Department of molecular genetics and biotechnology,
Yuriy Fedkovych Chernivtsi National University
2 Kotsiubynskoho str., 58012 Chernivtsi, Ukraine*

*During the evolution, plants have developed defense mechanisms against temperature stress associated with the synthesis of protective proteins, among which molecular chaperones predominate, in particular low molecular weight heat stress proteins (sHSPs). Despite the important role these proteins play in the defense response to high temperature stress, they are still poorly studied for many taxonomic groups of plants. In particular, almost nothing is known about the organization and diversity of heat shock genes/proteins in representatives of such an important genus as *Nicotiana* (tobacco). The article presents the results of a bioinformatic analysis of the multigene/multiprotein sHSP family in *N. sylvestris*, an important model object in plant physiology, biochemistry, molecular genetics and cell biotechnology.*

The GenBank database was used to search for homologous sequences using the BLAST algorithm. Amino acid sequences were aligned using the L-INS-I method on the MAFFT server. Phylogenetic analysis was performed by the maximum likelihood method using the PhyML plugin for Geneious Prime 2023.2.1. Statistical support of branches was calculated using the aLRT-Chi2 method.

*It was found that within the compared amino acid sequences of *N. sylvestris* sHSP, three parts can be distinguished: the variable N-terminal region, the conserved α -crystallin domain (ACD), and the C-terminal region. Based on the alignment of the sHSP amino acid sequences, a similarity dendrogram was constructed, on which several clades with high statistical support are visualized.*

*The results of the comparative analysis of the amino acid sequences show that sHSPs of *N. sylvestris* belong to 10 structural classes. Proteins of seven classes are believed to be localized in the cytoplasm and/or nucleus, while the rest are located in the endoplasmic reticulum, mitochondria, plastids and peroxisomes.*

*In total, according to the results of bioinformatic analysis, 24 genes encoding sHSP were found in the genome of *N. sylvestris*, as well as one pseudogene in which the fragment encoding the N-terminal region has been lost. Since the sequence of the pseudogene differs slightly from the closest representatives of this group, it can be assumed that the conversion of this region into a pseudogene is a relatively recent evolutionary event.*

*Keywords: abiotic stress, bioinformatic analysis, genetic polymorphism, molecular genomics, molecular evolution, multigene family, α -crystallin domain, sHSP, *N. sylvestris*.*

Отримано редколегією 02.05.2024 р.

ORCID ID

Людмила Козуб: <https://orcid.org/0000-0002-2675-5896>

Ірина Панчук: <https://orcid.org/0000-0002-2837-4480>