



ГЕНОМНІ ПІДХОДИ ДО МОНІТОРИНГУ МІКРОБІОМІВ ВЕТЛЕНДІВ ТА ОЗЕР

Стефані КАСТРО-МАРКЕС¹, Валтер ВОЛФСБЕРГЕР¹, Кеннет ГІЛКЕРТ¹, Ярослава ГАСИНЕЦЬ², Роман КІШ², Михайло ВАКЕРИЧ^{2,4}, Владислав МІРУТЕНКО², Олексій КУРУЦА², Андрій ВОЩЕПИНЕЦЬ², Віолета МІРУТЕНКО^{2,4}, Вікторія СТАНИНЕЦЬ², Одрі Дж. МАДЖЕСКЕ¹, Віктор ТОВТ³, Тарас ОЛЕКСИК^{1,2}

Екологічна метагеноміка змінює підходи до моніторингу водних об'єктів (озер, водосховищ, річок) і ветлендів шляхом прямого аналізу ДНК і РНК безпосередньо з води та донних відкладів. Мікроорганізми (в т.ч. бактерії, одноклітинні водорості, гриби) швидко реагують на екологічні, гідрологічні та лімнологічні зміни довкілля (швидкість течії, коливання рівня води, накопичення неорганічних і органічних речовин, температурні зсуви тощо), що дозволяє їм виступати індикаторами ранніх сигналів змін в екосистемі. Цей огляд присвячений практичним аспектам прісноводної метагеноміки бактеріопланктону, зокрема можливостям використання портативного нанопорового секвенування довгочитів (довгих рідів, long-read nanopore sequencing) для забезпечення моніторингу безпосередньо у польових умовах. У роботі узагальнено практичні варіанти моніторингу прісних водойм з акцентом на портативній технології секвенування Oxford Nanopore (ONT). Ця технологія дозволяє впроваджувати робочі процеси в режимі близькому до реального часу безпосередньо на місцях відбору проб, а також проводити повнорозмірне профілювання 16S рРНК для підвищення точності таксономічної ідентифікації. Порівняно дві взаємодоповнювальні стратегії: рутинне ампліконне профілювання, яке є дешевшим і зручним для частого відбору проб та шотган-метагеноміку (shotgun-metagenomics) – дорожчий, але більш інформативний для функціональної інтерпретації метод. Прагматична модель моніторингу полягає у використанні ампліконів для базового спостереження та переході до шотган (дробового, shotgun) або цільового функціонального аналізу, якщо проби відхиляються від базових показників. Інтерпретація структурована на трьох рівнях: (1) склад угруповання, (2) функціональний потенціал і (3) розуміння процесів та прогнозування. Варто наголосити, що надійність висновків залежить від плану-дизайну відбору проб, узгодженого із гідрологічною зв'язаністю і структурою оселищ, а також повторюваністю у просторі та часі. Оскільки 16S-таблиці зазвичай є композиційними (відносними), обговорюються статистичні практики, які зменшують ризик хибних сигналів (наприклад, уникнення рутинної розрідженості (rareфакції, rarefaction) та використання методів, що враховують композиційну природу даних, а також підкреслюється цінність додавання абсолютної шкали, де це можливо. Нарешті, на прикладі гідрологічного континуума на градації озеро-ветленд між оз. Балатон і прилеглим ветлендом Кіш-Балатон в Угорщині окреслено потенційні індикатори, які дозволяють відрізнити стабільний стан ядра мікробіому від локальних переходів та наведено короткий чек-лист звітування, що сприяє відтворюваності результатів та їх дозотривалому повторному використанню.

Ключові слова: екологічна ДНК, прісні води, водно-болотні угіддя, бактеріопланктон, довгочитне секвенування, Oxford Nanopore, 16S рРНК, екологічний моніторинг, бази даних.

¹Оклендський університет, Рочестер, Мічиган 48309, США; e-mail: oleksyk@oakland.edu

²Ужгородський національний університет, Ужгород, 88000, Україна; e-mail: yaroslava.hasynets@uzhnu.edu.ua

³Інститут лімнологічних досліджень Балатону HUN-REN, Тігань, 8237, Угорщина; e-mail: toth.viktor@blki.hu

⁴Закарпатський науково-дослідний експертно-криміналістичний центр Міністерства внутрішніх справ України, Слов'янська наб., 25, Ужгород, 88000, Україна

Кастро-Маркес С.: <https://orcid.org/0000-0002-9766-1299>

Волфсбергер В.: <https://orcid.org/0000-0003-0980-645X>

Гілкерт К.: <https://orcid.org/0000-0002-1530-0415>

Гасинець Я.: <https://orcid.org/0000-0003-4325-4695>

Кіш Р.: <https://orcid.org/0000-0002-7986-3372>

Вакерич М.: <https://orcid.org/0000-0002-3268-7797>

Мірутенко В.: <https://orcid.org/0000-0003-3845-5003>

Куруца О.: <https://orcid.org/0000-0001-7020-7996>

Вощепинець А.: <https://orcid.org/0000-0001-8307-781X>

Мірутенко В.: <https://orcid.org/0009-0002-9787-040X>

Станинець В.: <https://orcid.org/0009-0002-2670-1705>

Маджеске О.Дж.: <https://orcid.org/0000-0002-2178-5869>

Товт В.: <https://orcid.org/0000-0002-6182-2759>

Олексик Т.: <https://orcid.org/0000-0002-8148-3918>

Genomic approaches for monitoring wetland and lake microbiomes

Castro-Marquez S.¹, Wolfsberger W.¹, Hilkert K.¹, Hasynets Y.², Kish R.², Vakerych M.², Mirutenko V.², Kurutsa O.², Voshchepynets A.², Mirutenko V.^{2,4}, Stanynets V.², Majeske A.J.¹, Tóth V.³, Oleksyk T.^{1,2}

Environmental metagenomics is changing how we monitor water bodies (lakes, reservoirs, rivers) and wetlands by analyzing DNA and RNA directly from water and sediments. Microorganisms (algae, bacteria, fungi) respond quickly to environmental, hydrological, and limnological changes (water flow, level fluctuation, inorganic nutrients and organic matter accumulation, temperature shifts, etc.), allowing them to provide early signals of ecosystem change. This review focuses on practical freshwater metagenomics of bacterioplankton, especially how portable long-read nanopore sequencing can support real monitoring in the field. This review summarizes practical options for freshwater monitoring, with an emphasis on portable long-read Oxford Nanopore sequencing (ONT), which can support near-real-time, field-based workflows and full-length 16S rRNA profiling for improved taxonomic resolution. We contrast two complementary strategies: routine amplicon profiling (cost-effective for frequent sampling) and shotgun metagenomics (more resource-intensive but informative for functional inference). A pragmatic monitoring model uses amplicons for baseline surveillance and escalates to shotgun or targeted functional assays when samples deviate from baseline. We frame interpretation at three levels: (1) community composition, (2) functional potential, and (3) process understanding and forecasting. We also emphasize that reliable inference depends on sampling designs aligned with hydrologic connectivity and habitat structure, with replication across space and time. Because 16S tables are typically compositional, we discuss statistical practices that reduce false signals (e.g., avoiding routine rarefaction and using composition-aware methods), and we highlight the value of adding an absolute scale when feasible. Finally, using the Balaton–Kis-Balaton wetland–lake hydrological continuum as an applied example, we outline candidate indicators that separate stable “core” states from local transitions and provide a reporting checklist that supports reproducibility and long-term reuse.

Key words: environmental DNA, freshwater, wetlands, bacterioplankton, long-read sequencing, Oxford Nanopore, 16S rRNA, environmental monitoring.

¹ Oakland University, Rochester, MI 48309, USA; e-mail: oleksyk@oakland.edu

² Uzhhorod National University, 32, Uzhhorod, 88000 Ukraine; e-mail: yaroslava.hasynets@uzhnu.edu.ua

³ HUN-REN Balaton Limnological Research Institute, Tihany, 8237 Hungary; e-mail: toth.viktor@blki.hu

⁴ Transcarpathian scientific research expert and forensic center of the Ministry of Internal Affairs of Ukraine, 25, Slovianska nab., Uzhhorod, 88000, Ukraine

Castro-Marquez S.O.: <https://orcid.org/0000-0002-9766-1299>

Wolfsberger W.W.: <https://orcid.org/0000-0003-0980-645X>

Hilkert K.: <https://orcid.org/0000-0002-1530-0415>

Hasynets Y.: <https://orcid.org/0000-0003-4325-4695>

Kish R.: <https://orcid.org/0000-0002-7986-3372>

Vakerych M.: <https://orcid.org/0000-0002-3268-7797>

Mirutenko V.: <https://orcid.org/0000-0003-3845-5003>

Kurutsa O.: <https://orcid.org/0000-0001-7020-7996>

Voshchepynets A.: <https://orcid.org/0000-0001-8307-781X>

Mirutenko V.: <https://orcid.org/0009-0002-9787-040X>

Stanynets V.: <https://orcid.org/0009-0002-2670-1705>

Majeske A.J.: <https://orcid.org/0000-0002-2178-5869>

Tóth V.R.: <https://orcid.org/0000-0002-6182-2759>

Oleksyk T.K.: <https://orcid.org/0000-0002-8148-3918>

Вступ

Екологічна метагеноміка – це системне вивчення генетичного матеріалу, виділеного безпосередньо із зразків довкілля (наприклад, води, ґрунту, донних відкладів, повітря, біоплівки тощо). Її мета – описати, які організми присутні в середовищі, а також оцінити їхнє різноманіття, структуру угруповання та функціональний потенціал без виділення окремих видів у чисту культуру. Метагеномне секвенування забезпечує прямий, незалежний від культивування спосіб характеристики геномного «профілю» для будь-якого менеджменту мікроорганізмів.

Цей підхід полягає у виділенні ДНК та/або РНК безпосередньо з природних зразків і подальшому секвенуванні отриманих фрагментів. Далі прочитані послідовності порівнюють із референт-

ними базами даних, що містять анотовані геномні та таксономічні відомості, описані в попередніх дослідженнях і накопичені у базах даних онлайн. Таким чином, метагеномне секвенування дає змогу вивчати угруповання мікроорганізмів *in situ*, без культивування (без «агару в чашках Петрі») та створення контрольованих лабораторних систем, що особливо важливо для нових, ще не зафіксованих і рідкісних видів бактерій, або тих, які важко культивувати в лабораторних умовах.

Метагеномне секвенування давно вийшло за межі фундаментальної науки і широко використовується на практиці. Наприклад, у клініці воно підтримує діагностику інфекцій, швидке виявлення патогенів і нагляд за стійкістю бактеріальних збудників до антибіотиків. У природоохоронній діяльності метагеноміка допомагає від-

стежувати забруднення та пов'язані з ним зміни мікробних угруповань, а також оцінювати ефективність відновлювальних заходів. У екологічних дослідженнях ці методи стали основою для опису біорізноманіття, розуміння біогеохімічних циклів і моніторингу стану екосистем.

Одним із найвідоміших епізодів в історії метагеноміки є програма *Global Ocean Sampling* (GOS), пов'язана з морськими зборами проб (зразків) під час експедицій яхти *Sorcerer II* під егідою *J. Craig Venter Institute* (2004). Її мета полягала в тому, щоб застосувати шотган-метагеномне секвенування до ДНК, виділеної безпосередньо з морської води, і таким чином отримати прямий, незалежний від культивування зріз різноманіття та функціонального потенціалу океанічних угруповань мікроорганізмів (Venter et al. 2004). Показовою ранньою демонстрацією цього підходу стало дослідження Саргасового моря, яке продемонструвало, що секвенування екологічної ДНК відкриває доступ до величезного масиву раніше невідомих генів і ліній мікроорганізмів. Подальші дані та аналізи в межах GOS закріпили екологічну метагеноміку як дисципліну великомасштабного безпосереднього секвенування угруповань мікроорганізмів із навколишнього середовища, здатну виявляти просторову неоднорідність мікробів у природних екосистемах (Rusch et al. 2007). Саму експедицію навіть порівнювали з класичними морськими подорожами XIX століття, зокрема мандрівкою Дарвіна на кораблі *Бігль* (J. Craig Venter Institute 2004).

Ранні дослідження, що застосовували ампліфікацію гена малої субодиниці рибосомної РНК (SSU rРНК, small subunit ribosomal RNA) та клонові бібліотеки з природних оселищ переконливо показали переваги молекулярних підходів і показали, що ми не фіксуємо значну частину світу мікроорганізмів, коли покладаємося лише на вирощування в культурі (Amann et al. 1995; Hugenholtz et al. 1998). Таким чином, ключова перевага метагеноміки в екології – це здатність перетворювати об'єктивні дані, отримані шляхом секвенування, на прикладні екологічні висновки. Це дає змогу формувати індикатори стану, виявляти ранні попереджувальні сигнали змін і екологічних криз, а також розробляти практичні рекомендації для екологічного менеджменту. Серед оселищ, де такі підходи демонструють високу інформативність, особливо придатними є прісноводні озера: вони інтегрують вплив водозбору й часто чутливо реагують на антропогенні та кліматичні зміни, що робить їх зручними об'єктами для секвенування і аналізу на рівні угруповань.

У цьому огляді ми зосередимося на гідрологічно пов'язаних водно-болотних угіддях і прісноводних озерах, де мозаїчність середовища формує виражені градієнти складу та функцій угруповань мікроорганізмів. Метагеноміка озер – це секвенування ДНК та/або РНК безпосередньо зі зразків води, донних відкладів або біоплівки озера із метою опису угруповань мікроорганізмів і реконструкції їхнього функціонального потенціалу та ролей в екосистемі. Хоча лабораторні й аналітичні кроки багато в чому подібні до інших напрямів метагеноміки, мікробні угруповання озер мають властивості, що роблять їх особливо інформативними: вони швидко реагують на гідрологічні зміни, імпульсні надходження поживних речовин, динаміку розчиненої органічної речовини та сезонне перемішування води. Це робить озера перспективними системами для раннього виявлення і постійного моніторингу стану екосистеми (Newton et al. 2011). Метагеномний моніторинг є особливо привабливим для озер, оскільки він може бути неінвазивним і масштабованим, де навіть невеликі об'єми води дають багатий сигнал про угруповання; механістично інформативним, де зміни у таксономічному складі та функціональних профілях відображають ключові процеси екосистеми, а також дедалі придатнішим для довгострокового моніторингу бактеріопланктону в реальному часі завдяки портативним платформам секвенування і стандартизації аналітичних процесів (Quince et al. 2017; Urban et al. 2021).

У цьому огляді ми спираємося на наші власні дослідження озерної системи Центральної Європи, щоб узагальнити підходи та сучасні розробки екологічної метагеноміки й споріднених секвенувальних методів для моніторингу. Як приклад, ми розглядаємо гідрологічний континуум на градації озеро-ветленд – між оз. Балатон та прилеглим ветлендом Кіш Балатон в Угорщині (*Kis-Balaton Water Protection System* (KBWPS)), демонструючи, як портативне довгочитне (*long-read*) нанопорове (*nanopore*) секвенування повної довжини 16S рРНК ампліконів можна застосувати для високоякісного профілювання бактеріопланктону (Castro-Marquez et al. 2024). Такий підхід дає змогу одночасно описувати спільний мікробіом озера та виявляти закономірності вздовж переходу «озеро-ветленд». В підсумку ці молекулярні сигнали геномів присутніх у оселищі можна транслювати в екологічні індикатори, придатні для рутинного моніторингу й нагляду за станом прісноводних екосистем, із перспективою наближення до оперативних (майже реального часу) протоколів у польових умовах (Urban et al. 2021; Zhang et al. 2023).

1. Підходи і рівні метагеноміки

1.1. Ампліконне профілювання і шотган-метагеноміка

Ампліконне секвенування (*amplicon profiling*, або *metabarcoding*) – метод секвенування маркерних генів, зокрема 16S рРНК для бактерій та архей (ITS-регіону для грибів; а для ширших оглядів еукаріот – ділянок 18S рРНК), що забезпечує ефективну таксономічну інвентаризацію угруповань мікроорганізмів (Callahan et al. 2016; Bolyen et al. 2019). Ампліконне секвенування особливо придатне для досліджень із повторними відборами та високою просторово-часовою роздільністю, оскільки поєднує відносно низьку вартість із добре стандартизованими аналітичними робочими процесами (Bolyen et al. 2019). Історично цей підхід виріс із молекулярної філогенії та великих відкритих даних, що дозволили пряму ідентифікацію видового різноманіття мікроорганізмів без культивування (Woese, Fox 1977; Pace 1997).

Шотган-метагеноміка (*дробова, порційна метагеноміка, shotgun metagenomics*) навпаки

секвенує всю ДНК у пробі. Це дає змогу робити висновки не лише про таксономічний склад, а й про функціональний потенціал угруповання, виявляти віруси та реконструювати геноми нововиявлених організмів із суміші секвенованих фрагментів ДНК (*metagenome-assembled genomes, MAGs*). Водночас шотган-метагеноміка зазвичай є дорожчою і потребує складнішої обробки даних (Quince et al. 2017). Попри це, саме її здатність одночасно описувати таксономічну (“що тут є?”) та функціональну (“що воно тут робить?”) різноманітність зробила цей метод ключовим для перших великомасштабних екологічних досліджень і стимулювала запуск міжнародних ініціатив, таких як шотган-секвенування Саргасового моря та експедиції *Sorcerer II Global Ocean Sampling (GOS)*. У свою чергу, ці роботи створили великі каталоги нових генів і показали масштаб невідомого до цього часу різноманіття вірусів і мікроорганізмів у водних екосистемах (Venter et al. 2004; Rusch et al. 2007; Yooseph et al. 2007). Пізніше міжнародні

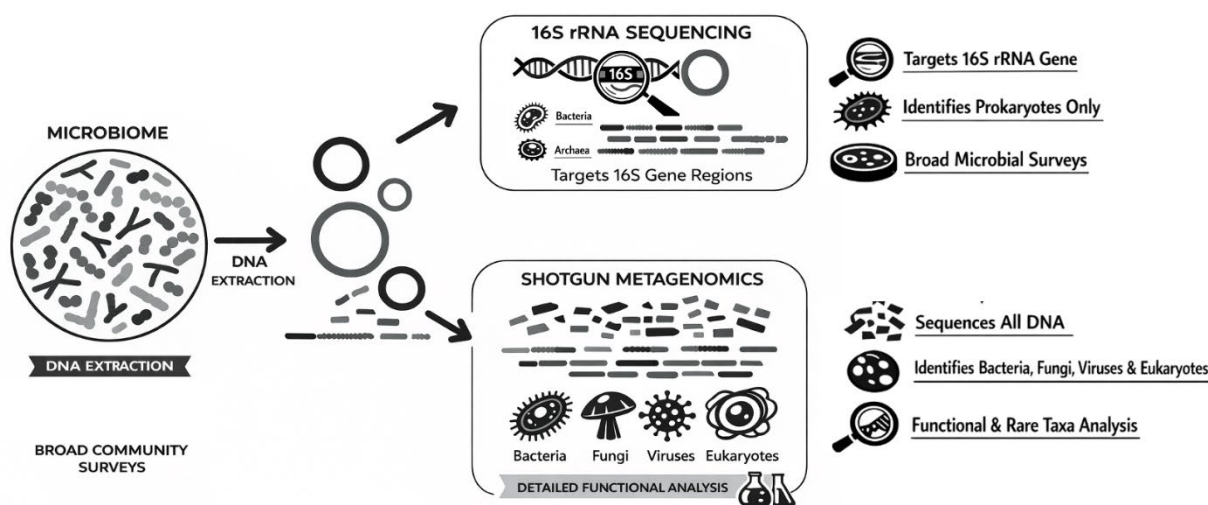


Рис. 1. Підходи метагеноміки: ампліконне профілювання і шотган-метагеноміка.

Fig. 1. Metagenomic approaches: amplicon profiling and shotgun metagenomics.

Примітка: (I) Секвенування гена 16S рРНК – це різновид ампліконного секвенування, який націлений на певні гіперваріабельні ділянки одного гена, що трапляється лише у бактерій та архей. (II) Шотган-метагеномне секвенування, навпаки, передбачає випадкове фрагментування всієї геномної ДНК у зразку, що дає змогу секвенувати всі ділянки ДНК, а не лише один маркерний ген. Ця принципова відмінність означає, що шотган-секвенування може одночасно виявляти бактерії, віруси, а також гриби та інші еукаріотичні мікроорганізми, тоді як стандартне 16S-секвенування обмежується прокаріотами. У підсумку, 16S-секвенування зазвичай рекомендують для широких оглядових описів складу угруповань мікроорганізмів, тоді як шотган-секвенування частіше обирають для детального функціонального аналізу або виявлення рідкісних таксонів.

Note: (I) Sequencing of the 16S rRNA gene is a form of amplicon sequencing that targets specific hypervariable regions of a single gene found exclusively in bacteria and archaea. In contrast, involves random fragmentation of all genomic DNA in a sample, enabling the sequencing of all DNA fragments rather than just a single marker gene. This fundamental difference means that shotgun sequencing can simultaneously detect bacteria, viruses, as well as fungi and other eukaryotic microorganisms, whereas standard 16S sequencing is limited to prokaryotes. Consequently, 16S sequencing is generally recommended for broad surveys of microbial community composition, while shotgun sequencing is more commonly used for detailed functional analyses or for detecting rare taxa.

ініціативи, зокрема *Tara Oceans*, розширили цю логіку до узгодженого відбору проб у глобальному масштабі з чітким екологічним контекстом, що прискорило перехід до опису функціональної організації угруповань і їхньої біогеографії (Sunagawa et al. 2015). У сучасній практиці шотган-метагеноміку найчастіше застосовують саме тоді, коли потрібно перейти від таксономічного опису до інтерпретації генетичного “інструментарію” мікробіоти: метаболічних шляхів, вірусного компоненту та генів, пов’язаних з ключовими екосистемними процесами.

У завданнях екологічного моніторингу дедалі частіше застосовують гібридну стратегію: регулярний скринінг відносно недорогим ампліконним секвенуванням для високочастотного моніторингу та відстеження динаміки угруповань, а у разі фіксації нетипової події ініціюють шотган-метагеномне секвенування для функціональної інтерпретації змін (Quince et al. 2017). Такий підхід є популярним, оскільки поєднує шотган-метагеноміку, яка дозволяє перейти від таксономічного опису до функціонального потенціалу, вірусного компоненту та ген аналізу геномів, зібраних із метагеномних даних (MAGs). У підсумку гібридний дизайн оптимізує витрати, частоту спостережень і глибину екологічних висновків.

Таблиця 1. Три рівні інтерпретації мікробіому: як метагеномні дані перетворюються на практичні результати для моніторингу та менеджменту.

Table 1. Three levels of microbiome interpretation: how metagenomic data are translated into practical outcomes for monitoring and management.

Рівень інтерпретації	Ключове питання	Виміри / аналіз	Результат моніторингу
1) Склад угруповання (community composition)	<i>Хто тут присутній?</i>	Таксони/ASV, різноманіття (α), відмінності між зразками (β), “ядро” мікробіому, індикаторні таксони	Карта/профіль угруповання за станціями і сезонами; порівняння між оселищами; виявлення відхилень від базового стану
2) Функціональний потенціал (functional potential)	<i>Що це угруповання може робити?</i>	Гени, функціональні категорії, метаболічні шляхи, функціональна різноманітність/ надлишковість; маркерні гени процесів (N/S/P/C, метан, фотосинтез тощо)	Оцінка “можливих” процесів у водоймі; порівняння функціональних профілів між станами; виявлення ризиків (напр., AMR/віруси — якщо релевантно)
3) Екологічні процеси (ecological processes)	<i>Чому угруповання змінюється і що буде далі?</i>	Часові ряди, повторні відбори, зв’язок із факторами оселища (t° , змішування, поживні речовини, каламутність); моделі/прогнози; невизначеність і валідація	Сигнали раннього попередження (цвітіння, гіпоксія); оцінка ризику; правила реагування (порогові критерії), що запускають додатковий відбір проб або дії менеджменту

1.2. Рівні інтерпретації мікробіому: склад, потенціал і процеси

У цьому розділі ми пропонуємо просту рамку для інтерпретації даних мікробіому при моніторинзі озер і водно-болотних угідь. На практиці метагеномні результати можна читати на трьох взаємопов’язаних рівнях: склад угруповання (хто присутній), функціональний потенціал (які функції та процеси спільнота потенційно здатна виконувати) та рівень процесів (чому угруповання змінюється і як ці зміни можна використати для прогнозування). Такий поділ допомагає перейти від опису до висновків, корисних для менеджменту: від “карти таксонів” до індикаторів ризику, сигналів раннього попередження та правил реагування. Таблиця 1 стисло підсумовує кожен рівень, ключові питання, типові показники та ті результати, які найчастіше потрібні для реального моніторингу.

Таблиця узагальнює три взаємопов’язані рівні аналізу: (1) склад угруповання, (2) функціональний потенціал і (3) екологічні процеси. Для кожного рівня наведено ключове питання, типові показники та приклади виходів, які найчастіше використовують у моніторингу озер і водно-болотних угідь (базові профілі, індикатори змін, оцінка ризику та правила реагування).

1.2.1. Рівень складу угруповання (*community composition*) описує таксономічний склад, структуру та різноманіття у пробах води або донних відкладів (Woese, Fox 1977; Pace 1997; Callahan et al. 2016; Bolyen et al. 2019). У водних системах ампліконне секвенування, у поєднанні з референсними базами даних і великими порівняльними проектами, підсиленими високопродуктивними секвенувальними платформами, перетворило опис складу мікробіому на відтворюваний екологічний сигнал, який можна кількісно характеризувати, картувати та порівнювати між оселищами, сезонами й роками.

На цьому рівні зазвичай використовують кілька груп характеристик: (I) показники внутрішньої різноманітності (α -diversity: багатство таксонів/ASV, індекси Шеннона та Сімпсона), (II) філогенетичні метрики (наприклад, філогенетична різноманітність (*phylogenetic diversity*), Faith's PD), (III) показники міжзразкових відмінностей (β -diversity: *Bray-Curtis*, *Jaccard*; філогенетичні відстані на кшталт *UniFrac*) із подальшим відображенням у просторі ординації (*PCoA/NMDS*) та тестуванням відмінностей між групами (наприклад, *PERMANOVA*), (IV) характеристики доміантності/рівномірності та стабільного ядра мікробіому (*core microbiome*), а також (V) підходи до виявлення таксонів-індикаторів і диференційної представленості між станами/біотопами (Bray, Curtis 1957; Faith 1992; Lozupone, Knight 2005; Anderson 2001).

1.2.2. Рівень функціонального потенціалу (*functional potential*) описує сукупність генів і функціональних ознак, що визначають на які біохімічні перетворення та екосистемні ролі угруповання мікроорганізмів потенційно здатне (тобто його метаболічні шляхи і функціональний арсенал). Найкраще цей рівень оцінювати за шотган-метагеномними даними, тоді як в окремих випадках (і з відповідними застереженнями) його також можна прогнозувати на основі секвенування ампліконів, але за умови достатнього покриття референсних геномів і коректної калібрації моделей (Langille et al. 2013; Quince et al. 2017; Douglas et al. 2020).

На цьому рівні функціональний потенціал зазвичай характеризують кількома групами показників: (I) геноцентричними профілями (таблиці відносної представленості функціональних категорій/родин генів та ортологів), (II) реконструкцією метаболічних шляхів і модулів (наявність/покриття ключових кроків, "повнота" шляхів, аль-

тернативні маршрути метаболізму), (III) функціональним різноманіттям і надлишковістю (скільки різних генетичних рішень підтримує одну й ту саму функцію; наскільки функції "розподілені" між таксонами), (IV) функціональною β -відмінністю між зразками (порівняння функціональних профілів між сезонами, біотопами, станами з подальшими ординаціями/тестами), (V) профілями цільових маркерних генів для ключових біогеохімічних процесів (цикли N/S/P/C, метаногенез/метанотрофія, фотосинтез/хемолітотрофія тощо), а також (VI) спеціалізованими категоріями на кшталт вірусного компоненту, генів антибіотикорезистентності, мобільних елементів і факторів взаємодії між організмами якщо вони релевантні для задач моніторингу та менеджменту.

Важливість функціонального потенціалу стала особливо очевидною після знакових шотган-досліджень довкілля, зокрема секвенування Саргасового моря та подальших глобальних оглядів океану, які виявили та продемонстрували масштаб новизни у оцінці генетичного різноманіття і закріпили метагеноміку як підхід до вивчення екосистем через функціональні профілі угруповань (Venter et al. 2004; Rusch et al. 2007; Sunagawa et al. 2015). Водночас широке застосування амплікон-профілювання сформувало потребу в акуратних рамках інтерпретації (наприклад, PICRUSt і наступні інструменти) та підкреслило ключове обмеження: точність прогнозів суттєво залежить від представленості відповідних таксонів і оселищ у референтних базах, тому для слабо охоплених екосистем похибка може бути значною (Langille et al. 2013; Douglas et al. 2020).

1.2.3. Рівень процесів показує перехід від опису до прогнозування. Тут ми не лише фіксуємо, які мікроорганізми присутні, а намагаємося зрозуміти, чому угруповання змінюється і що буде далі. Для цього ми поєднуємо три типи інформації: (1) динаміку угруповань мікроорганізмів у часі, (2) ключові чинники оселища (температура, змішування, надходження поживних речовин, каламутність тощо) і (3) механістичні обмеження, які визначають "швидкість" і напрям цих змін.

У контексті озер цей рівень усе частіше зводиться до прогнозу подій, важливих для менеджменту водойми: початку "цвітіння" води, розвитку гіпоксії (дефіциту кисню) та реакції екосистеми на відновлювальні заходи. Саме на цьому рівні повторні відбори проб і часові ряди перетворюють метагеномні патерни на сигнали раннього

попередження, які можна використовувати до того, як проблема стане очевидною за візуальними або хімічними показниками (Widder et al. 2016; Marrone et al. 2023).

Метагеномний моніторинг, довгі часові ряди та мультиомні підходи (*multi-omics*: наприклад, поєднання ДНК, РНК і метаданих про оселище) дають змогу зв'язувати зміни структури угруповань із реальними екосистемними процесами та підтримувати прогнозування (Faust et al. 2015; Linz et al. 2020; Marrone et al. 2023).

Оскільки прогнози в моніторингу завжди мають невизначеність, важливо подавати їх як ризик (ймовірність події) разом із простими довірчими інтервалами або діапазоном можливих значень, а не як “точне передбачення”. Щоб такі прогнози були надійними, їх потрібно перевіряти на попередніх даних (*back-testing*) і на незалежних сезонах або роках, порівнюючи передбачення з реальними спостереженнями. На практиці корисно заздалегідь визначити правила реагування або тригери: наприклад, якщо ризик цвітіння перевищує певний поріг, або якщо з'являються стабільні відхилення від базового “ядрового” угруповання, запускаємо додатковий відбір проб, уточнювальні аналізи або посилений контроль якості води. Такі правила перетворюють метагеномні сигнали на керований інструмент, де рішення приймаються раціонально, прозоро і повторювано. Водночас варто пам'ятати, що сильні кореляції з факторами середовища не завжди означають причинність, тому ключові висновки бажано підтверджувати незалежними вимірюваннями (наприклад, киснем, токсинами, пігментами або даними з супутників).

2. Переваги довгочитів (*long-reads*) у секвенуванні

Менш ніж двадцять років тому мікробиоту озер вивчали переважно непрямими методами. Озера часто сприймали як природні чашки Петрі, а зміни в угрупованнях оцінювали за лабораторно культивованими ізолятами, мікроскопією та біогеохімічними тестами. Але цей підхід мав фундаментальне обмеження: більшість мікроорганізмів, які можна розгледіти під мікроскопом у природних пробах, не ростуть на стандартних лабораторних поживних середовищах.

Молекулярні підходи змістили фокус із фенотипу на генотип і одночасно змінили як лабораторну частину моніторингу, так і обробку даних. Спочатку мікробиоту аналізували за допомогою клонів бібліотек і секвенування методом

Сенгера малих субодиниць рРНК. Цей підхід радикально розширив уявлення про біорізноманіття, але залишався відносно низькопродуктивним і тому обмеженим за кількістю проб (Pace 1997; Hugenholtz et al. 1998). Поява секвенування нового покоління (NGS), особливо короткочитного (*коротке прочитування, short-read*) секвенування Illumina, різко підвищила продуктивність. Зокрема, відносно недороге ампліконове 16S-секвенування дозволило аналізувати значно більше зразків, робити багато вимірів і повторів, що уможливило статистично потужні порівняння між оселищами і стаціями, басейнами, глибинами та сезонами (Caporaso et al. 2011 2012).

Однак для постійного та довготривалого моніторингу водних екосистем навіть із NGS залишаються дві ключові проблеми. По-перше, робочий процес часто прив'язаний до централізованих лабораторій: потрібен час на транспортування проб, підготовки бібліотек і отримання результатів. По-друге, короткочити мають обмежену довжину, і цього інколи недостатньо, щоб впевнено розрізняти близькі таксони та робити інтерпретацію на рівні виду.

Саме ці обмеження пояснюють, чому в метагеномі зростає роль портативного секвенування з довгочитами (довгими прочитуваннями, *long-read*) у реальному часі, зокрема високоточних довгочитових платформ на кшталт Oxford Nanopore та PacBio). Такий підхід змінює не лише те, скільки ми можемо секвенувати, але як швидко ми можемо відреагувати на екологічний сигнал. Нижче буде коротко показано, як довгочитове секвенування підсилює моніторинг завдяки портативності й потоку даних (*streaming*; 3.1), а також як повнодовжинний 16S покращує таксономічну роздільну здатність та екологічну інтерпретацію (3.2).

Хоча PacBio є надійною платформою довгочитів, зокрема завдяки *HiFi (high-fidelity)* секвенуванню, що покращує точність та підтримує високоякісні метагеномні реконструкції й функціональні висновки (Han et al. 2024; Kim et al. 2024), у наших прикладах ми зосереджуємося на *Oxford Nanopore* як на підході, що швидко розвивається та вже доступний в Україні (Oleksyk et al. 2025).

2.1. Портативність і оперативність

Пристрій *MinION* від *Oxford Nanopore Technologies (ONT)* дає змогу виконувати портативне секвенування і отримувати потік даних (*streaming*). Результати секвенування стають доступними практично одразу, під час роботи приладу, дозволяючи моніторинг у реальному часі

(*real-time*). Перші мобільні лабораторії *ONT* показали, що навіть мінімального польового набору достатньо, щоб секвенувати ДНК у віддалених і екстремальних умовах. Наприклад, у тропічному лісі секвенування *ONT* застосовували для швидкого ДНК-баркодингу та перевірки зразків на місці без надійного електропостачання (Menegon et al. 2017). В Сухих долинах Антарктиди *ONT* секвенування залишалося працездатним попри труднощі польової роботи та обмежену інфраструктуру (Johnson et al. 2017). Паралельно *ONT* стає інструментом оперативного геномного нагляду під час спалахів інфекційних хвороб. Найвідоміший приклад застосування відбувся під час епідемії еболи в Західній Африці, коли фахівці не чекали транспортування матеріалу до віддалених лабораторій секвенування, обробляючи зразки на місці, в польовому госпіталі (Quick et al. 2016). Працездатність технології перевіряли навіть під час космічних польотів, що підкреслює її надійність поза звичними лабораторними умовами (Castro-Wallace et al. 2017).

У випадку прісноводних екосистем, ранні дослідження вже продемонстрували, що профілювання угруповань мікроорганізмів та швидка діагностика мікробіомів озер і річок можуть виконуватися у форматі та з термінами, які реально вписуються в програми моніторингу (Urban et al. 2021). Ключова перевага, це мінімальна затримка між відбором проби й отриманням даних: коли секвенування проводять у полі або одразу після відбору, зменшується ризик післявідборових змін у складі угруповання та деградації ДНК. Секвенування *ONT* проходить в онлайн-режимі, а тому дає результат майже одразу, тому його можна використати, щоб вчасно відреагувати (доповнити зборами проб, перевірити ризик, запустити уточнювальні тести), а не лише пояснювати події після завершення.

Таким чином, для моніторингу водних систем, головні переваги технології *ONT*, це портативність і оперативність. Секвенування можна виконувати під час польової кампанії, що дозволяє швидко уточнювати план відбору проб, виконувати адаптивний повторний відбір, робити висновки, рекомендації і попередження поєднуючи дані з лімонологічними показниками і заздалегідь визначеними пороговими критеріями і тригерами.

2.2. Роздільна здатність

Ключова технічна відмінність між довгочитними та короткочитними платформами (*long-read vs. short read*) – це довжина прочитувань і саме

вона визначає роздільну здатність на практиці. Короткочитні 16S-дослідження зазвичай націлені на одну або дві гіперваріабельні ділянки (наприклад, V4 або V3–V4). Вони стали основою ранньої епохи вивчення мікроорганізмів. У той самий час стало очевидним обмеження короткочитних платформ: вибір праймерів і ампліфікованого регіону може змінювати таксономічні визначення, тим більше, ще близькоспоріднені таксони можуть мати майже однакові послідовності коротких фрагментів (Caporaso et al. 2011; Johnson et al. 2019).

На відміну від короткочитних платформ типу Illumina, технологія *ONT* може секвенувати повнодовжинні або майже повнодовжинні амплікони 16S рРНК (~1,4-1.6 кб). Такі прочитування охоплюють значно більше інформативних ділянок, ніж короткі фрагменти з однієї гіперваріабельної зони. Це зазвичай підвищує таксономічну роздільну здатність, особливо коли для класифікації використовують регулярно оновлювані референсні бази, наприклад SILVA (Quast et al. 2013).

Для моніторингу озер і водно-болотних угідь це дуже важливо, бо багато “попереджувальних” ознак пов’язані з конкретними групами мікроорганізмів. Це можуть бути мікроби, які з’являються або різко зростають під час цвітіння води, групи, споріднені з фекальними індикаторами або прісноводні лінії, що виконують різні екологічні ролі. Часто такі групи відрізняються дуже тонко – на рівні роду або виду. Якщо секвенувати лише короткий фрагмент, цієї інформації може не вистачити, і замість точних таксонів ми отримуємо надто широкі та нечіткі таксономічні категорії.

Емпіричні дослідження показують, що повнодовжинний 16S часто дозволяє визначати бактерії до рівня виду, а інколи – робити висновки навіть на рівні штамів, якщо аналітичні пайплайни та референсні бази даних достатньо якісні й добре підібрані (Johnson et al. 2019). Новіші результати для нанопорних платформ із хімією R10.4.1 також свідчать про вищу точність на рівні нуклеотидів і краще видове профілювання при роботі з повнодовжинним 16S (Zhang et al. 2023). Для задач моніторингу прісних вод це має як наукове, так і практичне значення: дослідження показують, що нанопорний 16S дає екологічно інформативні профілі угруповань, які можна інтегрувати у рутинні робочі процеси польового або лабораторного моніторингу (Urban et al. 2021).

Окремо варто згадати підходи з унікальними молекулярними ідентифікаторами (UMI). Якщо “помітити” вихідні молекули ДНК до ПЛР і потім зібрати консенсус-послідовності, можна істотно

зменшити помилки окремих прочитувань. Це послаблює компроміс між точністю та масштабом/швидкістю для майже повнодовжинного 16S на нанопорних системах. Такий підхід особливо корисний для довгострокових програм моніторингу, де критично важлива порівнюваність результатів між роками та різними гідрологічними режимами (Lin et al. 2024).

Матеріали та методика

3. Проектування моніторингу з нанопорним секвенуванням

У прісноводних озерах і водно-болотних угіддях мікроорганізми швидко реагують на рух води, перемішування та межі між різними оселищами. Це означає, що результат аналізу сильно залежить від того, де і коли взято пробу. Водночас технічні кроки роботи з ДНК, насамперед фільтрація та екстракція в каламутній воді з великою кількістю частинок, можуть суттєво впливати на те, які саме мікроорганізми потраплять у зразок і наскільки надійно їх можна буде виміряти. Тому відбір проб, який ігнорує гідрологію, може пропустити справжню структуру екосистеми або переплутати сезонний шум із важливими екологічними змінами (Knight et al. 2012), а невдалий підхід до отримання ДНК може створювати системні помилки (*bias*), які виглядають як еко-

логічні відмінності, але насправді є наслідком протоколу або умов підготовки проб. Саме тому етап проектування моніторингу є визначальним для того чи стане озерна метагеноміка надійним інструментом дослідження, або залишиться разовим спостереженням.

У цьому розділі ми зосереджуємося на двох ключових складових моніторингової метагеноміки для озер і водно-болотних угідь: (1) як використати гідрологію, щоб правильно визначити, де і коли відбирати проби, і (2) як підібрати стратегії фільтрації та екстракції (разом із регулярними контролями), щоб біологічні патерни відображали процеси що відбуваються безпосередньо у водоймі, а не побічні лабораторні ефекти. Таблиця 1 підсумовує мінімальний список робіт для 16S-моніторингу ONT у водно-болотних угіддях та озерах, а наступний розділ коротко пояснює кожен пункт.

3.1. Дизайн відбору проб

У прісноводних озерах і ветлендах угруповання мікроорганізмів найчастіше формуються насамперед гідрологією. На практиці це означає, що режим перемішування, час перебування (*residence time*) води, зв'язок між басейнами та притоками, а також межі між різними оселищами (водне дзеркало біля прибережної зони, ділянки з макрофітами, можливі шари водної товщі) визна-

Таблиця 2. Мінімальний список для 16S-моніторингу з допомогою ONT у водно-болотних угіддях та озерах.
Table 2. Minimum checklist for 16S monitoring using ONT in wetlands and lakes.

Підхід	Що потрібно описати / що це дає	Як це інтерпретувати (навіщо це потрібно)
1. Дизайн відбору проб	Показати, як точки відбору пов'язані з рухом води (притоки/витоки, басейни). Вказати, скільки повторів зібрано і коли саме (час доби, сезон) (Knight et al. 2012).	Робить результати порівнюваними та допомагає зрозуміти, чи можуть патерни бути пов'язані з часом/місцем відбору, а не з реальними змінами в екосистемі.
2. Фільтрація та екстракція	Розмір(и) пор мембрани, який об'єм води фільтрували, як працювали з інгібіторами, і чи націлювалися на вільноживучі мікроорганізми, прикріплені до частинок, або на обидві групи.	Вибір фільтра може змінювати, які мікроорганізми ви “зловите”, особливо в каламутних болотах і мілких озерах.
3. Контролі	Перелік польових “порожніх” проб (field blanks), “порожніх” екстракцій (extraction blanks) і модельних/штучних угруповань (mock communities). Вказати, що означає “контроль пройшов”, і що робити, якщо контроль не проходить (наприклад, перезапуск, виключення партії).	Допомагає виявити контамінацію, інгібування та проблеми пайплайну до того, як вони виглядатимуть як “справжні” екологічні зміни.
4. Метадані секвенування	Тип проточної комірки (flowcell), реагенти (наприклад, R10.4.1), версія/ модель basecaller, налаштування демультіплексування.	Зміни в програмі або реагентів

чають рамки, на яких вибудовується склад мікробіому. Оскільки ці чинники працюють у масштабах від окремих штормів до багаторічних змін, дизайн моніторингу має бути побудований так, щоб відрізнити справжні екологічні сигнали від сезонного “шуму” та просторової автокореляції. Саме тому потрібні повтори у часі та відбір проб уздовж гідрологічних і екологічних градієнтів, а також чітке документування того, де і коли взято кожен зразок (Knight et al. 2012).

Підтвердження того, що гідрологія структурує угруповання мікроорганізмів, наведені для різних типів озер і регіонів. Наприклад, у північних озерах Європи зміни часу перебування води пов’язані зі зсувами типових прісноводних бактеріальних груп. Це узгоджується з ідеєю, що системи з коротким обігом води у більшій мірі визначаються притоками з водозбору та так званими мас-ефектами (ситуацією, коли організми постійно заносяться ззовні) і ця дисперсія помітно впливає на склад угруповання). У озерах помірної зони окремі епізоди перемішування діють як екологічні “порушення”: вони можуть швидко перебудувати бактеріальні угруповання, а їхня інтенсивність і сезонність впливають на те, як угруповання розходяться між ділянками та як швидко повертаються до попередніх станів.

Довготривалі, повторювані спостереження додатково показують, що тип озера та режим перемішування можуть впливати на склад і динаміку мікробіому. Зокрема, багаторічні дослідження демонструють повторювану структуру угруповань боліт і озер, яка водночас реагує на сезонні та щорічні фізичні зміни (Linz et al. 2020). У великих евтрофних мілких системах гідрологія взаємодіє з градієнтами поживних речовин формуючи просторові патерни. Наприклад, в озері Тайху (Китай) склад бактеріопланктону змінюється в просторі й часі вздовж трофічних градієнтів, що підкреслює важливість врахування як водного руху, так і локальних умов середовища. Бактеріальні метаболіти в мережах річок і озер лісових зон півночі (тобто у взаємопов’язаних річках, струмках і озерах) часто організуються під впливом пов’язаності водойм і напрямку течії. Таксони можуть переноситися вниз за течією, але локальні умови “фільтрують”, хто з них зможе закріпитися й домінувати на конкретних ділянках.

Оскільки рух води та перемішування формують угруповання мікроорганізмів, дизайн відбору проб має охоплювати головні гідрологічні осі системи (притоки/витоки, басейни, ділянки з різним часом перебування води, межі між біо-

топами) і бути повтореним у часі. У звітуванні важливо показати, як станції відбору пов’язані з рухом води, а також вказати кількість повторів і точні часові рамки (час доби, сезон), тому що саме ці деталі роблять результати порівнюваними між сезонами та роками і дозволяють впевненіше пов’язати зміни у профілях секвенування з реальними змінами екосистеми, а не з випадковим вибором місця чи дати (Knight et al. 2012).

3.2. Фільтрація проб і виділення ДНК

У водній товщі завжди співіснують принаймні дві великі фракції: вільноживучі клітини та мікроорганізми, прикріплені до частинок (органічний детрит, глина, водорості, агрегати). Те, яку частину угруповання можна буде побачити у результатах, напряму залежить від того, який фільтр ви обрали, який об’єм води реально вдалося профільтрувати, як ви працювали з інгібіторами, і на яку фракцію ви націлювалися (вільноживучу, прикріплену до частинок або обидві). Таким чином, фільтрація та екстракція ДНК є одним із тих етапів, на яких необхідно запобігати появі технічних артефактів.

У болотах і мілких озерах ця проблема заго-струється через високе навантаження частинками, яке сильно змінюється у просторі та часі. Наприклад, під час вітрового підняття донного осаду, після злив і стоку з водозбору або на піку біомаси фітопланктону підвищується каламутність води і тому схема фільтрації, яка добре працює у прозорій відкритій воді, може дуже швидко зіпсуватися в каналі з густим ростом макрофітів або в прибережній “плямі” (*plume*), де вода несе багато завислих частинок.

Коли фільтр забивається, це не просто незручність і втрата часу. Забивання може змістити спектр захоплення у бік більших по розміру частинок, зменшувати профільтрований об’єм і, що критично, концентрувати разом із ДНК речовини-інгібітори. У природних пробах такими інгібіторами часто стають продукти розкладання органіки, пігменти, метали та інші домішки, які впливають на протікання ПЛР і заважають у підготовці бібліотек. У результаті цих відхилень під час секвенування, склад угруповання може непомітно спотворюватися: частина таксонів втрачається, а асоційовані на частинках угруповання, що переважають, виглядають більш важливими ніж насправді.

Саме тому сучасні водні протоколи дедалі частіше орієнтуються не на єдине універсальне рішення, а на підбір методу фільтрації під конкретний тип середовища та цільову фракцію (вільножи-

вучі клітини vs. асоційовані на частинках мікроорганізми). На практиці це може означати поетапну фільтрацію або використання різних розмірів пор, щоб контролювано відокремити фракції та зменшити технічні перекося у складі угруповання; відомо, що сам вибір пористості фільтра здатен змінювати видимий профіль бактеріальних угруповань у прісній воді (Vyarranahalli et al. 2021). Так само екстракцію ДНК часто потрібно пристосувати до каламутних або органічно насичених вод: додати етапи очищення, подовжити лізис і застосувати спеціальні кроки для видалення інгібіторів ПЛР (Schrader et al. 2012). Щоб такі рішення залишалися стандартними та порівнюваними між біотопами й сезонами, важливо не лише підібрати правильний протокол, а й системно документувати ключові параметри (тип фільтра/пори, об'єм, умови виділення, контроль якості) як частину стандартного опису мета-даних (Yilmaz et al. 2011). В цілому, краще не оптимізувати кожен сайт окремо, а забезпечити результати відтворюваними й порівнюваними в межах усієї системи (Goldberg et al. 2016).

3.3. Контрольні вимірювання

У практиці моніторингу мінімальний набір зазвичай включає негативні контролю, тобто порожні проби (*field blanks*), фільтри, екстракції (*extraction blanks*) та позитивні контролю, наприклад модельні угруповання (*mock communities*). Порожні проби дозволяють побачити контамінацію з повітря, тари, або з рук під час відбору; бланки фільтрації та екстракції показують, чи не додає сигнал власне сама підготовка проб; позитивні контролю перевіряють, чи система взагалі працює так, як очікується, і чи зберігається точність між серіями проб.

Мета контрольних вимірювань – відокремити реальні зміни у водоймі від технічних артефактів, які можуть виникати під час відбору, фільтрації, екстракції, ПЛР або біоінформатичної обробки. Без системи контролів контамінація, інгібітори або зміни в пайплайні можуть імітувати екологічний сигнал, спотворюючи моніторинг реальних змін. Нарешті, у звітуванні важливо фіксувати не лише факт їх наявності, а й правила прийняття рішень: пороги/критерії прийнятності і що саме робиться у разі якщо контроль не дав бажаного результату (перезапуск секвенування, повторна екстракція, виключення партії, повторний відбір тощо). У підсумку регулярні контролю дають головну практичну перевагу: коли в даних з'являється відхилення, у вас є підстава сказати, що

зміна ймовірно сталася у водоймі, а не виникла через контамінацію, інгібіцію, чи помилку у пайплайні. Саме це робить метагеномні дані придатними для довготривалого нагляду й порівняння між сезонами та роками.

3.4. Мета-дані секвенування

У програмі моніторингу головна мета полягає в забезпеченні порівнянності результатів між різними локаціями, сезонами та роками. Це означає, що біоінформатичний аналіз теж має бути відтворюваним, а не одноразовим запуском, який неможливо повторити. Оскільки нанопорові дані залежать від вибору реагентів та програмного забезпечення (зокрема *basecalling*), правил фільтрації, якості, способу демультіплексування і версій довідкових баз, ці параметри потрібно документувати щоразу.

У звіті для ампліконового секвенування 16S варто фіксувати щонайменше: тип слайду (*flow cell type*), хімію (наприклад, R10.4.1), версію набору (*kit*), версію та модель програмного забезпечення (*basecaller*), а також налаштування демультіплексування. Зміни в наборах, слайдах, реагентах або програмному забезпеченні можуть впливати на якість прочитувань, ідентифікацію і визначення таксонів, що проявляються як партійні (серійні) ефекти (*batch effects*). Тобто навіть якщо водойма не змінилося, може бути зафіксований інший (хибний) склад угруповання і далі помилково інтерпретований як екологічна зміна. Без мета-даних такі відхилення легко переплутати з реальними трендами в екосистемі. Саме тому метадані секвенування настільки важливі: вони є частиною вимірювання, що дозволяє відрізнити технічні аберації від екологічних змін.

Практично, ці вимоги підтримують інструменти для відтворюваної біоінформатики. QIIME 2 важливий тим, що зберігає походження, джерело: автоматично записує дані, параметри і кроки аналізу, тож повторивши алгоритм схеми дослідження (*workflow*) можна отримати той самий результат (Bolyen et al. 2019). Системи керування робочими процесами, такі як Snakemake і Nextflow, перетворюють аналіз на чіткий рецепт, який стабільно запускається на різних комп'ютерах або інфраструктурах (Köster, Rahmann 2012; Di Tommaso et al. 2017). А групові ініціативи на кшталт nf-core допомагають зменшити хаотичні відмінності між командами та роками, підтримуючи пайплайни в перевіреному, контрольованому стані (Ewels et al. 2020). У контексті моніторингу це означає просту

річ: якщо помічено зміну в даних секвенування – є підстави вважати, що вона пов’язана з водою, а не з останнім оновленням інструментів.

3.5. Обробка даних секвенування

Обробка даних секвенування є етапом, на якому сирі послідовності (raw reads), отримані з нанопорового секвенатора, проходять фільтрацію та таксономічну класифікацію і зрештою перетворюються на фінальну таблицю складу мікробіому. Для довгочитного 16S таксономічну класифікацію зазвичай виконують двома основними шляхами. Перший: вирівнювання (alignment), тобто пряме зіставлення прочитуваних із референсними послідовностями; для цього часто використовують *minimap2* (Li 2018). Другий: швидкі класифікатори для оперативного скринінгу (в реальному часі). Такі інструменти, як *Centrifuge* або *Kraken 2*, швидко класифікують прочитування за *k-mer*-підходом і дають змогу швидко оцінити, чи виглядає проба незвичною відносно базового стану ділянки (Kim et al. 2016; Wood et al. 2019). Водночас результати цих методів істотно залежать від того, яку саме базу даних використано і коли її оновлювали, тому ці параметри потрібно чітко фіксувати, щоб тренди можна було коректно порівнювати між сезонами та роками.

Окремо важливо занотувати, як саме виконували демультіплексування та обрізування (trimming) праймерів/адаптерів (інструмент і ключові параметри), бо помилки на цьому кроці можуть призводити до плутанини прочитуваних між зразками. Химерні послідовності (артефакти ПЛР) потрібно виявляти й видаляти на етапі обробки, щоб вони не створювали фальшиві нові таксони та не спотворювали оцінку складу й різноманіття. Кількість химер зростає при високій кількості циклів ПЛР/низькій ДНК концентрації, тому вміст химер може відрізнятися між сезонами. Так само потрібно чітко визначити формат фінального результату: це може бути таблиця кількостей прочитуваних, згрупованих за ASV/OTU, або агрегована таксономічна таблиця (рід/вид) із прив’язкою до конкретної версії/релізу референсної бази для класифікації (наприклад, SILVA) (Quast et al. 2013).

3.6. Статистика

Більшість статистичних таблиць 16S-моніторингу мікробіому описують відносні частки, а не абсолютну кількість клітин у воді. Тобто ми бачимо, який відсоток прочитуваних належить кожному таксону, але не знаємо кількості біомаси чи клітин у літрі води. Через це мікро-

біомні дані називають композиційними: якщо частка одних таксонів зростає, частка інших має зменшитися, навіть коли загальна кількість мікроорганізмів в об’ємі не змінилася. Якщо трактувати такі відносні дані як абсолютні, можна отримати неправильні висновки та хибні кореляції (Gloor et al. 2017; Quinn et al. 2019).

З практичної точки, для моніторингу краще застосовувати методи, які працюють саме з композиційною природою даних: лог-відношення (*log-ratios*) та їхні варіанти (наприклад, CLR/ILR-підходи) замість відносних часток або прямих порівнянь кількості прочитаних послідовностей між зразками (Quinn et al. 2019). Інтерпретацію значно підсилює прив’язка до абсолютної шкали: наприклад, qPCR, проточна цитометрія або spike-in стандарти, які дозволяють оцінювати не лише зміну частин, але й зміну загального мікробного навантаження (Vandeputte et al. 2017). Для моніторингу водойм це особливо корисно, коли цікавлять не просто домінуючі таксони, а зміна загальної біомаси/кількості клітин, як у випадках з ризиком гіпоксії, цвітіння тощо.

У ампліконних 16S-даних часто виникає технічна нерівність: різні проби мають різну кількість прочитаних послідовностей (різну “глибину” секвенування). Через це може з’являтися спокуса “вирівняти” дані підвибіркою (*rarefying*) – тобто випадково відібрати і залишити однакову кількість прочитуваних у кожній пробі. На перший погляд це здається інтуїтивним підходом, але на практиці він має суттєвий недолік: навмисно відкидається частина інформації, а разом із нею, частина статистичної достовірності (*statistical power*). Це особливо відчутно, коли глибина секвенування значно відрізняється між пробами. У крайніх випадках доводиться або обрізати до найгіршої проби (тобто проби з найменшою кількістю прочитаних послідовностей ДНК), або втрачати цілі зразки, які не дотягують до порогу. У результаті порівняння між місцями, сезонами чи роками можуть стати менш стабільними, а деякі відмінності – виглядати відчутнішими або слабшими лише через те, як саме виконали підвибірку. Саме тому для більшості моніторингових задач рекомендують не обрізати (*trim*) дані випадковою підвибіркою, а застосовувати методи, які прямо враховують різну глибину секвенування у статистичній моделі або нормалізації, тобто контролювати кількість прочитаних послідовностей як технічний фактор, а не прибирати послідовності з найкращих наборів даних (McMurdie, Holmes 2014; Weiss et al. 2017).

Різні методи можуть давати різні списки (диференційна представленість/чисельність) важливих таксонів навіть на тих самих даних. Систематичні дослідження показують, що узгодженість між інструментами не є гарантованою, тому для моніторингу бажано робити перевірки чутливості: чи зберігається головний висновок, якщо застосувати 2–3 коректні методи та пороги чутливості (Nearing et al. 2022; Weiss et al. 2017). Оскільки дані композиційні, диференційні тести слід робити методами, які це враховують: ANCOM та ALDEx2 (Mandal et al. 2015; Fernandes et al. 2014). У звіті моніторингу доцільно чітко зазначати, який метод використано, які пороги чутливості/налаштування застосовано, і чи узгоджуються ключові сигнали між методами.

Для довготривалих програм (сезони/роки) з'являється ще одна типова проблема: серійний ефект (ефект партії, batch effect), коли відмінності між серіями виділення ДНК, бібліотек, слайдів (*flowcell*), версій програмного забезпечення (*basecaller*) або навіть різними днями роботи лабораторії можуть створювати артефакти даних. Для мікробіомних досліджень підходи до корекції/контролю серійного ефекту розглядають як важливий елемент коректного висновку. Якщо його не контролювати, можна переплутати технічну зміну з екологічним трендом. Практична рекомендація для аналізу довготривалих програм, щоб виключити серійний ефект – рандомізувати проби між серіями, додавати технічні коваріати у моделі (серія/партія/слайд/версія програмного забезпечення *basecalling*) та регулярно запускати стабільний контроль, щоб бачити стандартну похибку інструмента.

Нарешті, моніторинг майже завжди означає багато тестів одночасно (сотні таксонів, десятки порівнянь). Тому статистичні результати треба трактувати разом із контролем помилок множинного тестування (*multiple testing*). Найчастіше це робиться через визначення частки хибних відкриттів (*false discovery rate* (FDR); Benjamini, Hochberg 1995), що визначає, яка частка вірогідних результатів може бути хибною, коли тестів багато. Процедура Benjamini–Hochberg робить це, впорядковуючи *p-value* і встановлюючи поріг, який відповідає заданому рівню FDR. Але для менеджменту водойми важливо не лише значення *p*, а й розмір ефекту, його стабільність у часі та узгодженість із незалежними вимірюваннями (кисень, хлорофіл, токсини, каламутність). У моніторинговій практиці статистично значущий результат є лише частиною картини: рішення мають спиратися на поєднання статистики, екологічної інтерпретації

та контрольованої якості даних (Weiss et al. 2017; Nearing et al. 2022).

3.7. Доступність даних і коду

Довготривалий метагеномний моніторинг корисний лише настільки, наскільки його можна перевірити, відтворити та продовжити через рік або через п'ять років, навіть якщо зміниться дослідник, програміст чи програмне забезпечення. Якщо метагеноміка претендує на роль інструмента моніторингу, тоді дані, код і метадані стають частиною самого вимірювання, так само важливою, як фільтрація, контроль чи дизайн відбору. Тому мінімальна вимога моніторингу – відкриті бази даних прочитаних послідовностей (*raw reads*) у надійному публічному репозиторії, наданих разом із застосованим робочим алгоритмом схеми дослідження (*workflow*) з вказаними версією та параметрами, щоб аналіз можна було повторити (Bolyen et al. 2019). У цьому розділі буде обґрунтовано стандарти зі звітування та відтворюваності 16S-моніторингу озер і водно-болотних угідь (Таблиця 3).

Одних баз даних прочитаних послідовностей і моніторингу недостатньо без контексту. Щоб інші дослідники могли коректно порівнювати результати між сезонами, роками та локаціями, потрібні повні метадані: де і коли взято пробу, який тип середовища, які вимірювання зроблено в полі та які ключові параметри пробопідготовки використано. Саме для цього існують стандарти на кшталт MIxS/MIMARKS, які задають мінімальний набір полів, що робить дані придатними для повторного аналізу й порівняння між дослідницькими групами (Yilmaz et al. 2011). Також моніторинговим програмам допомагає управління даними у стилі FAIR (щоб дані було легко знайти та повторно використовувати з часом) і стабільний спосіб розповсюдження програм (наприклад, Bioconda), щоб зменшити проблему “працює лише на моєму комп'ютері”, коли змінюються інструменти або персонал (Grüning et al. 2018; Wilkinson et al. 2016).

Корисна паралель для екологічного моніторингу, консорціум MetaSUB (Metagenomics and Metadesign of Subways and Urban Biomes). Це глобальна співпраця, яка розглядає міста (особливо системи громадського транспорту) як стандартизовані “платформи відбору проб” для метагеноміки. MetaSUB показав, що узгоджений, багатсайтовий відбір проб зі спільними протоколами та багатими метаданими може створювати порівнювані “базові лінії” мікробіому між локаціями, але

Таблиця 3. Практичний список для звітування та відтворюваності в нанопорному 16S-моніторингу боліт і озер.
Table 3. Practical checklist for reporting and reproducibility in nanopore 16S monitoring of wetlands and lakes.

№	Стандарт	Мінімальні кроки	Потреба
1	Сирі дані	Завантажити <i>raw reads</i> у відкритий репозиторій та в тексті вказати <i>accession/</i> ідентифікатор набору даних.	Щоб інші могли перевірити результати, повторити обробку та використати дані в майбутніх порівняннях.
2	Пайплайн	Дати посилання на репозиторій/архів із <i>workflow</i> , вказати версію/реліз, ключові параметри та версії баз даних (Bolyen et al. 2019).	Щоб аналіз можна було відтворити без “прихованих” налаштувань і порівнювати між командами та роками.
3	Метадані	Заповнювати ключові поля за MlXs/MIMARKS (де/коли/як зібрано пробу, основні параметри оселища, протоколи тощо). (Yilmaz et al. 2011)	Без метаданих дані важко інтерпретувати: навіть якісне секвенування стає “сиротою” без контексту.
4	Відтворюваність	Зафіксувати, як встановлювали залежності (наприклад, через <i>Bioconda</i>), і які версії пакетів використовували, щоб мінімізувати партійні ефекти (<i>batch effects</i>) через різні середовища (Grüning et al. 2018).	Щоб уникати ситуації “працює лише на моєму комп’ютері” і технічних розбіжностей між запусками.
5	Порівнюваність	У кожному сезоні/році зазначати, чи змінювалися версії інструментів і референсних баз; якщо так – коротко пояснити, як це враховано в аналізі (Wilkinson et al. 2016).	Щоб не переплутати реальні екологічні зміни з технічними зсувами через оновлення софту/баз даних.

Примітка: Мінімальні кроки, які потрібно описати у звіті такі, щоб дані та результати довготривалого ампліконного моніторингу були перевірюваними й порівнюваними між сезонами та роками: депонування сирих прочитувань (*accession*), публікація версійованого алгоритму схеми дослідження (*workflow*) із параметрами та базами даних, стандартизовані метадані (MlXs/MIMARKS), фіксація програмного середовища для відтворюваності (наприклад, *Bioconda*) та документування змін версій інструментів/референсів у часі (Bolyen et al. 2019; Grüning et al. 2018; Wilkinson et al. 2016; Yilmaz et al. 2011).

Note: This table outlines the minimum steps that should be described in reports to ensure that data and results from long-term amplicon monitoring are verifiable and comparable across seasons and years. These include: deposition of raw sequencing reads in a public repository (with accession numbers); publication of a versioned analytical workflow including parameters and reference databases; use of standardized metadata (MlXs/MIMARKS); documentation of the software environment to ensure reproducibility (e.g., via *Bioconda*); and recording changes in tool and reference database versions over time (Bolyen et al. 2019; Grüning et al. 2018; Wilkinson et al. 2016; Yilmaz et al. 2011).

водночас зберігати локальні сигнатури та сигнали, важливі для громадського здоров’я (Afshinneko et al. 2016). Для водойм та озер немає еквіваленту MetaSUB, радше є набір комплементарних платформ і стандартів, які разом дають той самий ефект: порівнюваність між місцями та роками. Далі приведемо приклади деяких з існуючих платформ.

По-перше, це мережі, що стандартизують моніторинг середовища і дизайн спостережень. Наприклад, мережа GLEON (Global Lake Ecological Observatory Network) виросла саме з потреби робити порівнюваними високочастотні дані сенсорів (температура, кисень, стратифікація тощо) між озерами, щоб підтримувати моделі та менеджерські рішення. Для метагеноміки урок GLEON дуже прямий: якщо фізика й хімія води збираються за правилами, то біологічні (чи мікроорганізмів) часові ряди стають набагато легше інтерпретованими, бо їх можна прив’язати до

режимів перемішування, притоків, сезонів і подій (Hamilton et al. 2015).

По-друге, це інфраструктури, що стандартизують протоколи відбору/метадані на великій території. Тут показовий приклад мережа NEON у США. Стандартизація для водойм – це насамперед стандарти метаданих і протоколів, а не лише вибір секвенатора (National Ecological Observatory Network [NEON] 2024). Тому перевага NEON в тому, що вона нав’язує дисципліну протоколів, метаданих і довгострокового планування, без яких моніторинг із часом перетворюється на набір несумісних проектів.

По-третє, це екосистема стандартів для власне молекулярних даних. MlXs/MIMARKS задають мінімальний набір того, що треба описати для маркерних послідовностей, щоб чужі дані можна було зрозуміти й коректно порівняти (Yilmaz et al. 2011). У водних програмах це практично означає: навіть якщо сайти оселищ різні (болото vs озеро),

дослідник все одно прив'язує дані до однакового формату опису проби, середовища, протоколу, платформи і версій.

По-чверте, це глобальні “парасолькові” ініціативи, які не є водними за фокусом, але дають уніфіковані протоколи та уявлення про масштаб. Тому кожен моніторинг може розглядатися як частина більшого глобального руху до сумісних екологічних даних, а не як “локальну екзотику”. Earth Microbiome Project показав, як стандартизація протоколів і метаданих дозволяє збирати й порівнювати дані між дуже різними екосистемами і саме сумісність даних відкриває шлях до багатомасштабних висновків (Thompson et al. 2017). Ocean Sampling Day показав приклад узгодженого глобального відбору зразків у один часовий момент і зі спільними правилами, що робить дані порівнюваними між локаціями (Korf et al. 2015).

Щоб підтримати висновки “моніторингового рівня”, метагеноміка озер і ветлендів спирається на спільний інструментарій, який охоплює три пов'язані потреби: відтворювану обробку, обґрунтоване таксономічне призначення та статистику, що враховує властивості ампліконних даних. Практична рамка для системи “озеро-ветленд” найкраще виглядає як комбінація: (I) сенсорного/гідрологічного каркаса в стилі GLEON, (II) дисципліни протоколів і метаданих у стилі NEON + MIXS/MIMARKS та (III) відкритої, FAIR-орієнтованої публікації даних і коду, щоб часові ряди не втрачали сенс з часом (Wilkinson et al. 2016).

Приклад: моніторинг континууму озера Балатон – ветленд Кіш-Балатон

Озеро Балатон (46°50' N, 17°40' E; Рис. 2) – мілководне прісноводне озеро на заході Угорщини та найбільше в Центральній Європі за площею поверхні. Озеро має витягнуту форму: середня ширина близько 7,6 км, а максимальна – приблизно 14 км. Хоча озеро значне за площею (~596 км²), його глибина невелика (середня глибина близько 3,5 м; максимальна ~11 м), тому фізичні процеси тут швидко впливають на стан води і на те, що ми побачимо в біологічних даних. Північний берег озера оточений пагорбами передгір'я Баконських гір з вулканічними ґрунтами і відомий ландшафтами з виноградниками, тоді як південний берег переважно рівнинний і значною мірою зайнятий рекреаційною інфраструктурою (UNESCO 2017).

За типом змішування оз. Балатон є поліміктичним: у ньому зазвичай не формується стабільна

сезонна стратифікація, а під час штормів дрібний донний осад легко підіймається у товщу води. Через це вода часто виглядає каламутною, з високою концентрацією неорганічних зважених частинок (Istvánovics et al. 2007). Для моніторингу це принципово: і гідрологія, і перемішування можуть змінювати локальні умови за короткий час, а отже дизайн відбору проб має відображати просторову структуру озера.



Рис. 2. Загальний вигляд озера Балатон з півострову Тигонь (Tihany).

Fig. 2. General view of Balaton Lake from the Tihany Peninsula.

З практичної точки зору, оз. Балатон зручно систематично описувати через його поділ на чотири основні басейни уздовж гідрологічних градієнтів із заходу на схід (Рис. 3). Такий поділ важливий як робоча схема: басейни відрізняються впливом притоків, перемішуванням і часом утримання води, а отже можуть мати різні “підписи” мікробіому.

Ключовою частиною цього прикладу є система водного захисту Кіш-Балатону (KBWPS). Колишнє озеро, а сьогодні – напівштучний комплекс водно-болотних угідь, створений для перехоплення поживних речовин, зважених частинок та органічних речовин з вод річки Зала (Zala) до того, як ці води потраплять в озеро Балатон (Tátrai et al. 2000). Для екологічної геноміки ця система особливо цінна тим, що формує чіткий гідрологічний континуум річка–ветленд–озеро. Тобто є можливість спостерігати, як угруповання мікроорганізмів змінюються між пов'язаними оселищами вздовж реального шляху потоку води (Рис. 3).

Результати та обговорення

Повно-довжинне 16S-профілювання на платформі ONT уздовж континууму Балатон–KBWPS дає два типи сигналів, які особливо зручні для моніторингу. По-перше, всередині KBWPS видно

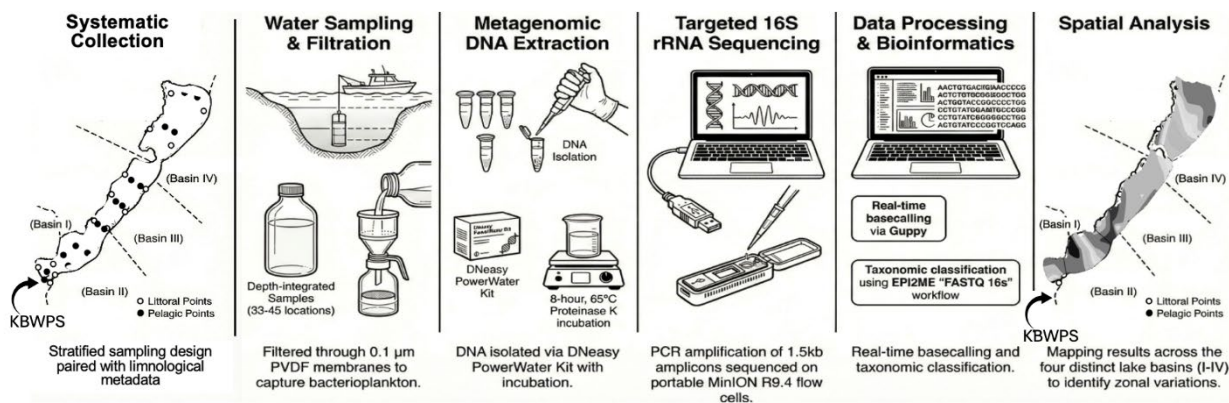


Рис. 3. Моніторинг мікробіому озера Балатон з урахуванням гідрологічного контексту та польових досліджень (модифіковано з Castro-Marquez et al. 2025).

Fig. 3. Microbiome monitoring of Lake Balaton incorporating hydrological context and field-based analysis (modified from Castro-Marquez et al. 2025).

Примітка: (I) систематичний підхід до відбору проб для моніторингу озера уздовж гідрологічних градієнтів (II) набір проб та фільтрацію; (III) виділення метабеномної ДНК; (IV) нанопорове секвенування повнодовжинного 16S у режимі потокового аналізу; (V) обробка та біоінформатичний дата аналіз даних та біоінформатика; (VI) карта з результатами просторового аналізу моніторингу.

Note: (I) a systematic sampling strategy for lake monitoring along hydrological gradients; (II) sample collection and filtration; (III) extraction of metagenomic DNA; (IV) nanopore sequencing of full-length 16S rRNA genes using real-time streaming analysis; (V) data processing and bioinformatic analysis; and (VI) a map presenting the spatial monitoring results.

різкі та повторювані переходи угруповань між з'єднаними секторами ветленду, тобто ветленд працює як “ланцюжок різних середовищ”, а не як однорідна зона. По-друге, в самому озері Балатон простежується спільне ядро бактеріопланктонного угруповання, яке повторюється між басейнами, але поверх нього накладається просторова структура, пов'язана з басейнами та впливом притоків (Castro-Marquez et al. 2025; Tóth et al. 2026). Така картина узгоджується з загальною ідеєю типових ліній бактеріопланктону, які повторюються в різних прісноводних озерах, але реагують на режим перемішування, поживні речовини та властивості органічної речовини (Newton et al. 2011).

Система Балатон–KBWPS також добре підходить для моніторингового дизайну тому, що її стан змінюється під впливом гідрологічного регулювання, кліматичної мінливості та змін у надходженні речовин із водозбору. Це означає, що станції та повторення потрібно планувати так, щоб вони були прив'язані до гідрології та часу (пори року), інакше можна не відрізнити реальні зміни від ефектів відбору проб (Hatvani et al. 2011; 2020; Kovács et al. 2012). У цьому сенсі континуум річка–ветленд–озеро можна розглядати як природний експеримент: він дозволяє перевіряти, як змінюються угруповання вздовж пов'язаних оселищ і як ці зміни пов'язані з вимірними екологічними градієнтами (Farkas et al. 2022; Tóth 2023).

Практично було застосовано два взаємодоповнювальні дизайни. Для оз. Балатон було створено стратифіковану мережу станцій за басейнами та ключовими типами оселищ (наприклад, прибережна зона проти відкритої води), щоб фіксувати градієнти, пов'язані з притоками, перемішуванням і часом утримання води (Castro-Marquez et al. 2025). Для KBWPS, закладено трансекту вздовж шляху водного потоку від річки Зала через сектори ветленду до їхнього витоку та потрапляння до водного басейну озера, щоб зробити зв'язок між водоймами “віссю”, яку можна порівнювати між сезонами й роками (Tóth et al. 2026). В обох випадках молекулярні проби супроводжувалися лімнологічними вимірюваннями, щоб інтерпретація мікробних зсувів була пов'язана з конкретними, вимірними умовами середовища (Castro-Marquez et al. 2025; Tóth et al. 2026). Такий підхід відповідає загальним рекомендаціям для відтвореного метабеномного моніторингу, де дизайн і повторення є ключем до надійних висновків (Knight et al. 2012).

У підсумку цей кейс показує, як портативну 16S-метабеноміку з ONT можна перевести з рівня разового опису в інструмент постійного моніторингу. У оз. Балатон ми отримуємо базову лінію у вигляді спільного “ядра” бактеріопланктону та просторову структуру між басейнами, а у KBWPS – чіткі переходи угруповань між

пов'язаними оселищами ветленду. Разом це дає дві комплементарні метрики для нагляду: стабільність ядра і швидкість/масштаб заміни угруповань (*turnover*) як потенційно чутливий індикатор змін стану системи, релевантних для менеджменту водних ресурсів (Castro-Marquez et al. 2025; Tóth et al. 2026).

Висновки

Екологічна метагеноміка у внутрішніх водах швидко переходить від простого опису таксонів у системі до підходу рівня нагляду, який може підтримувати своєчасні менеджерські рішення. Один із головних рушіїв цього процесу – портативне секвенування Oxford Nanopore (ONT), яке дозволяє отримувати дані безпосередньо в полі та працювати з ними в режимі потоку (*streaming*), а не очікувати тижні на результати з централізованої лабораторії (Urban et al. 2021). Друга важлива перевага – це можливість працювати з довшими ампліконами, включно з повнодовжинним 16S рРНК (~1,45–1,55 кб), що зазвичай дає кращу таксономічну роздільну здатність, ніж короткі фрагменти однієї гіперваріабельної ділянки (Zhang et al. 2023).

Сама по собі передова технологія не гарантує якості висновків. Щоб метагеномні сигнали справді відображали стан водойми, їх потрібно вбудувати у правильний дизайн спостережень. Для озер і ветлендів це особливо важливо, бо мікроорганізми швидко реагують на рух води, переміщення, притоки/витоки та межі між оселищами. Якщо дизайн відбору проб ігнорує гідрологію, зростає ризик переплутати просторові або сезонні коливання з “ новою проблемою”, якої насправді немає (Knight et al. 2012).

З практичної точки зору менеджерам зазвичай потрібні не “повні списки таксонів”, а зрозумілі індикатори, які можна пов'язати з рішеннями. Наприклад: наскільки стабільною є базове (ядро) угруповання; чи з'являються аномальні групи, яких раніше не було; чи пов'язані мікробні зсуви зі змінами екологічних змінних, таких як температура, каламутність, хлорофіл-а або CDOM. Саме тут перевага ONT у часі стає критичною: коли дані приходять одразу, їх можна використати для оперативної дії (додатковий відбір, уточнювальні вимірювання, посилення контролю), а не лише для ретроспективного звіту наприкінці сезону (Urban et al. 2021).

Окрема умова стабільного моніторингу – відтворюваність. Тому додаткові метадані – це важлива частина моніторингу (Bolyen et al. 2019).

У нанопорових даних результат помітно залежить від версій реагентів слайду/*flowcell*, *basecalling*-моделі, параметрів фільтрації якості та від того, яку саме референсну базу й реліз використано для таксономічного призначення. У довгих часових рядах це прямо зменшує ризик переплутати похибки процесу секвенування з реальними екологічними змінами.

Не менш важливою є статистична інтерпретація. Ампліконні таблиці 16S зазвичай відображають відносні частки, а не абсолютну кількість клітин. Через це дані мають композиційну природу, і “наївні” кореляції можуть бути хибними: якщо частка однієї групи зростає, частка іншої неминуче падає, навіть без змін у реальній біомасі (Gloor et al. 2017). Там, де це можливо, корисно додавати абсолютну шкалу (qPCR, точна цитометрія, *spike-in* стандарти), щоб відрізнити “перерозподіл часток” від реальних змін загального мікробного навантаження (Vandeputte et al. 2017). Так само варто обережно ставитися до підвибірки (*rarefying*): вона здається простим способом вирівняти глибину секвенування, але фактично відкидає дані й часто зменшує потужність статистичного аналізу; у моніторингу зазвичай краще застосовувати методи, що статистично враховують різну глибину, не “обрізаючи” інформацію (McMurdie, Holmes 2014; Weiss et al. 2017). Нарешті, різні методи диференційної представленості можуть давати різні списки таксонів, тому корисно робити перевірки чутливості та прозора звітувати, наскільки стабільні висновки (Nearing et al. 2022).

І, нарешті, довготривалий моніторинг має сенс лише тоді, коли його можна перевірити й продовжити. Мінімальний стандарт – відкритість даних, чіткі супроводжуючі метадані та програми аналізу з анотованими параметрами, щоб їх можна було легко знайти, отримати, інтерпретувати й повторно використати (Wilkinson et al. 2016). У підсумку це перетворює метагеноміку на інструмент, який реально підтримує моніторинг і менеджмент водними екосистемами.

У підсумку ми подаємо основні біоінформаційні фреймворки та аналітичні методи, які найчастіше використовують для цього: від систем робочих процесів і платформ, що стабілізують обробку з року в рік (Bolyen et al. 2019; Di Tommaso et al. 2017; Köster, Rahmann 2012), до курованих референсних ресурсів та інструментів для вирівнювання/класифікації довгочитів, які визначають, як саме маркуються прочитування (Kim et al. 2016; Li 2018; Quast et al. 2013; Wood et al. 2019).

Таблиця 4 також підкреслює статистичні підходи, що враховують композиційну природу 16S-даних і заохочують перевірки чутливості, включно з фреймворками композиційного аналізу та інструментами диференційної чисельності, поширеними в дослідженнях мікробіомів (Fernandes et al. 2014; Gloor et al. 2017; Mandal et al. 2015; Nearing et al. 2022; Weiss et al. 2017).

Рекомендації органам екологічного менеджменту для метагеномного моніторингу водно-болотних угідь та прісноводних озер

Ми пропонуємо, щоб комплексні стратегії менеджменту озер інтегрували екологічну метагеноміку в рутинний нагляд. Органам управління варто впроваджувати ієрархічну систему моніторингу, яка використовує економічно ефективне ампліконне профілювання для частих “переписів” (високочастотного спостереження), а шотган-метагеноміку або метатранскриптоміку залишає для механістичних досліджень, які запускаються у відповідь на аномальні екологічні події. Щоб оптимізувати діагностику майже в реальному часі,

Таблиця 4. Інструменти, стандарти та статистичні підходи для відтворюваного 16S-довгочитового моніторингу.

Table 4. Tools, standards, and statistical approaches for reproducible 16S long-read monitoring.

Роль у моніторингу	Визначення	Приклади	Мінімальна звітність	Потреба
1. Стандарти опису даних і контексту	Правила, які “пояснюють” пробу та роблять дані порівнюваними	MIxS/MIMARKS (Yilmaz et al. 2011); FAIR (Wilkinson et al. 2016)	Поля метаданих (де/коли/як відібрано; параметри середовища); де зберігаються дані; умови повторного використання	Щоб через 2–5 років дані можна було правильно інтерпретувати й порівняти
2. Відтворюваність обчислень (workflow)	“Рецепт” аналізу, який можна повторити	Snakemake (Köster, Rahmann 2012); Nextflow (Di Tommaso et al. 2017); nf-core (Ewels et al. 2020)	Посилання на workflow; версія/реліз; параметри; контрольні файли; як запускали (локально/НРС/хмара)	Щоб аналіз не залежав від “людини, яка тоді запускала”
3. Платформа аналізу ампліконів	“Комбайн” для стандартних кроків 16S-аналізу з provenance	QIIME 2 (Bolyen et al. 2019)	Версія QIIME 2; плагіни/кроки; налаштування; експортовані артефакти/звіт	Щоб результати були відтворювані й простежувані
4. Середовище програм (залежності)	Спосіб стабільно встановлювати однакові версії програм	Bioconda (Grüning et al. 2018)	Як керували версіями (environment.yml/lockfile); ключові версії інструментів	Щоб зменшити різницю через різні “програмні середовища”
5. Референсні бази	Довідники, від яких залежить таксономія	SILVA (Quast et al. 2013)	Назва бази + реліз/версія; як будували/фільтрували базу	Щоб оновлення не маскувалося як екологічний тренд
6. Таксономічне визначення (алайнмент/класифікація)	Інструменти, які перетворюють прочитування на таксони	minimap2 (Li 2018); Kraken 2 (Wood et al. 2019); Centrifuge (Kim et al. 2016)	Який метод; пороги якості; версія бази; параметри класифікації	Різні підходи можуть давати різні профілі навіть на тих самих даних
7. Статистична інтерпретація 16S-таблиць	Робота з композиційністю (відносні дані)	Композиційний підхід (Gloor et al. 2017; Quinn et al. 2019)	Які трансформації/метрики; як контролювали глибину секвенування	Щоб уникати хибних кореляцій і “ефектів частки”
8. Диференційна представленість (DA)	Методи, які порівнюють “що змінилось” між станами	ANCOM (Mandal et al. 2015); ALDEx2 (Fernandes et al. 2014)	Який метод; пороги; чи перевіряли надійність висновків	Різні DA-методи можуть давати різні списки таксонів
9. Перевірки чутливості та бенчмаркінг	Практика: показати, що висновки не “ламкі”	Порівняння методів/стабільність результатів (Nearing et al. 2022; Weiss et al. 2017)	Чи зберігаються ключові висновки при зміні методу/порогів	Особливо важливо для моніторингу, де рішення приймають за трендами
10. НЕ робити “за замовчуванням”	Методологічне попередження	Rarefying як default (McMurdie, Holmes 2014)	Якщо застосовували : чому; які наслідки для потужності	Це може “обрізати” дані й послабити висновки

Примітка: компоненти класифіковані за їхню роллю в моніторинговому процесі: (1) стандарти опису даних, (2) відтворюваний запуск аналізу, (3) таксономічне визначення/референси, (4) статистична інтерпретація та перевірки чутливості. Ці кроки зменшують “змішування рівнів” і полегшують відтворюване щорічне звітування (Bolyen et al. 2019; Di Tommaso et al. 2017; Köster, Rahmann 2012; Quast et al. 2013; Wilkinson et al. 2016; Yilmaz et al. 2011).

Note: components are categorized by their role in the monitoring process: (1) data description standards, (2) reproducible analysis run, (3) taxonomic definition/references, (4) statistical interpretation and sensitivity checks. These steps reduce “level mixing” and facilitate reproducible annual reporting (Bolyen et al. 2019; Di Tommaso et al. 2017; Köster, Rahmann 2012; Quast et al. 2013; Wilkinson et al. 2016; Yilmaz et al. 2011).

політикам і регуляторам слід підтримувати адаптивні, напівавтоматичні протоколи відбору проб, які надають пріоритет секвенуванню генетичних маркерів, важливих для менеджменту. Сюди належать маркери, пов'язані зі шкідливими “цвітіннями” водоростей, індикаторами фекального забруднення або антибіотикорезистентності.

Ефективні програми моніторингу мають бути побудовані з урахуванням гідрологічних характеристик цільової системи, щоб точки відбору відображали типові водні маси та різноманіття біотопів, зокрема відкритоводні басейни, літоральні зони та водно-болотні угіддя з домінуванням макрофітів. Менеджерським органам слід застосовувати градієнтні дизайни відбору проб, які враховують гідрологію й лімнологію керованих водойм, а також зв'язність між притоками та витоками. Збір даних потрібно підтримувати з частотою, достатньою для опису надійної сезонної динаміки й формування базових ліній; під час значних екологічних збурень варто планувати інтенсивніші серії відбору проб.

Технічні протоколи екстракції ДНК і фільтрації слід адаптувати до умов конкретної ділянки та ретельно документувати всі особливості. Практикам варто відходити від простих оцінок відносної чисельності й переходити до кількісного профілювання мікробіому, додаючи методи абсолютного масштабування – наприклад, кількісну ПЛР (qPCR), проточну цитометрію або внутрішні spike-in стандарти.

Політичні ініціативи мають надавати пріоритет створенню й підтримці керованих регіональних і локальних довідкових баз даних, специфічних для прісноводних екосистем, щоб максимально використати та розширити таксономічну роздільну здатність, яку дають повнодовжинні послідовності. Майбутні менеджерські втручання повинні прагнути міждисциплінарної інтеграції: поєднувати індикатори угруповань мікроорганізмів із гідродинамічними чисельними моделями

та продуктами супутникового дистанційного зондування, щоб перейти від описових оцінок до прогнозування переходів станів екосистем. Розглядаючи зміну складу угруповань (*turnover*) як чутливий індикатор “фільтрації” середовищем і функціонального здоров'я системи, регулятори зможуть краще оцінювати ефективність систем водного захисту та завчасно передбачати ризики, зокрема гіпоксію або початок “цвітіння”.

Подяки

Це дослідження виконувалося в межах магістерської роботи (MS Thesis) з біологічних наук і було підтримане стартовим фондом Оклендського університету та програмою гостьового стажування в Балатонському лімнологічному інституті Центру екологічних досліджень Угорської академії наук.

Фінансування

Дослідження Віктора Р. Товта були підтримані проектом RRF-2.3.1-21-2022-00008 та стипендією Fulbright 2022. Стефані О. Кастро-Маркес, Валтер В. Волфсбергер і Тарас Х. Олексик отримали підтримку для проведення досліджень через стартовий фонд Оклендського університету, а роботи Тараса Х. Олексика також були виконані в рамках гостьового стажування у Балатонському лімнологічному інституті Центру екологічних досліджень Угорської академії наук.

Конфлікт інтересів

Автори заявляють про відсутність конфлікту інтересів.

Дозволи

У цьому дослідженні не проводилися роботи за участю тварин або людей, воно не проводилося на приватних чи охоронюваних територіях. Для відповідних локацій не вимагалось спеціальних дозволів.

AFSHINNEKOO, E., MEYDAN, C., CHOWDHURY, S., JAROUDI, D., BOYER, C., BERNSTEIN, N., MARITZ, J. M., REEVES, D., GANDARA, J., CHHANGAWALA, S., AHSANUDDIN, S., SIMMONS, A., NESSEL, L., SUNDARESH, B., PEREIRA, E., JØRGENSEN, T. S., KOLOKOTRONIS, S.-O., KIRCHBERGER, N., GARCIA, I., ... MASON, C. E. (2016) Geospatial resolution of human and bacterial diversity with city-scale metagenomics. *Microbiome*, 4, 24. <https://doi.org/10.1186/s40168-016-0168-z>

AMANN, R. I., LUDWIG, W., SCHLEIFER, K.-H. (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiological Reviews*, 59(1), 143–169. <https://doi.org/10.1128/mr.59.1.143-169.1995>

ANDERSON, M. J. (2001) A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, 26(1), 32–46. <https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x>

BENJAMINI, Y., HOCHBERG, Y. (1995) Controlling

- the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57, 289-300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- BOLYEN, E., RIDEOUT, J. R., DILLON, M. R., BOKULICH, N. A., ABNET, C. C., AL-GHALITH, G. A., ALEXANDER, H., ALM, E. J., ARUMUGAM, M., ASNICAR, F., BAL, Y., BISANZ, J. E., BITTINGER, K., BREJNROD, A., BRISLAWN, C. J., BROWN, C. T., CALLAHAN, B. J., CARABALLO-RODRÍGUEZ, A. M., CHASE, J., ... CAPORASO, J. G. (2019) Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, 37(8), 852–857. <https://doi.org/10.1038/s41587-019-0209-9>
- BRAY, J. R., CURTIS, J. T. (1957) An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs*, 27(4), 325–349. <https://doi.org/10.2307/1942268>
- BYAPPANAHALLI, M. N., NEVERS, M. B., KORAJKIC, A., WHITMAN, R. L. (2021) Influence of filter pore size on bacterial communities in Lake Michigan coastal waters. *Frontiers in Microbiology*, 12, 665664. <https://doi.org/10.3389/fmicb.2021.665664>
- CALLAHAN, B. J., MCMURDIE, P. J., ROSEN, M. J., HAN, A. W., JOHNSON, A. J. A., HOLMES, S. P. (2016) DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13(7), 581–583. <https://doi.org/10.1038/nmeth.3869>
- CAPORASO, J. G., LAUBER, C. L., WALTERS, W. A., BERG-LYONS, D., LOZUPONE, C. A., TURNBAUGH, P. J., FIERER, N., KNIGHT, R. (2011) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences*, 108 (Supplement 1), 4516–4522. <https://doi.org/10.1073/pnas.1000080107>
- CASTRO-MARQUEZ, S. O., TÓTH, V. R., KOLCHANOVA, S., WOLFSBERGER, W. W., OLEKSYK, T. K. (2025) A long-read sequencing approach to high-resolution profiling of bacterioplankton diversity in a shallow freshwater lake. *Scientific Reports*, 15, 96558. <https://doi.org/10.1038/s41598-025-96558-7>
- CASTRO-WALLACE, S. L., CHIU, C. Y., JOHN, K. K., STAHL, S. E., RUBINS, K. H., MCINTYRE, A. B. R., ... BURTON, A. S. (2017) Nanopore DNA sequencing and genome assembly on the International Space Station. *Scientific Reports*, 7, 18022. <https://doi.org/10.1038/s41598-017-18364-0>
- DI TOMMASO, P., CHATZOU, M., FLODEN, E. W., BARJA, P. P., PALUMBO, E., NOTREDAME, C. (2017) Nextflow enables reproducible computational workflows. *Nature Biotechnology*, 35(4), 316–319. <https://doi.org/10.1038/nbt.3820>
- DOUGLAS, G. M., MAFFEI, V. J., ZANEVELD, J. R., YURGEL, S. N., BROWN, J. R., TAYLOR, C. M., ... LANGILLE, M. G. I. (2020) PICRUSt2 for prediction of metagenome functions. *Nature Biotechnology*, 38(6), 685–688. <https://doi.org/10.1038/s41587-020-0548-6>
- EWELS, P. A., PELTZER, A., FILLINGER, S., PATEL, H., ALNEBERG, J., WILM, A., GARCIA, M. U., DI TOMMASO, P., NAHNSEN, S. (2020) The nf-core framework for community-curated bioinformatics pipelines. *Nature Biotechnology*, 38(3), 276–278. <https://doi.org/10.1038/s41587-020-0439-x>
- FAITH, D. P. (1992) Conservation evaluation and phylogenetic diversity. *Biological Conservation*, 61(1), 1–10. [https://doi.org/10.1016/0006-3207\(92\)91201-3](https://doi.org/10.1016/0006-3207(92)91201-3)
- FARKAS, M., SZOBOSZLAY, S., VÖRÖS, L., LOVÁSZ, Z., MÉHES, N., MÁTYÁS, K., ... KRISZT, B. (2022) Bacterial community dynamics along a river–wetland–lake system. *Water*, 14(21), 3519. <https://doi.org/10.3390/w14213519>
- FAUST, K., LAHTI, L., GONZE, D., DE VOS, W. M., RAES, J. (2015) Metagenomics meets time series analysis: Unraveling microbial community dynamics. *Current Opinion in Microbiology*, 25, 56–66. <https://doi.org/10.1016/j.mib.2015.04.004>
- FERNANDES, A. D., REID, J. N., MACKLAIM, J. M., MCMURROUGH, T. A., EDGELL, D. R., GLOOR, G. B. (2014) Unifying the analysis of high-throughput sequencing datasets: Characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. *Microbiome*, 2, 15. <https://doi.org/10.1186/2049-2618-2-15>
- GLOOR, G. B., MACKLAIM, J. M., PAWLOWSKY-GLAHN, V., EGOZCUE, J. J. (2017) Microbiome datasets are compositional: And this is not optional. *Frontiers in Microbiology*, 8, 2224. <https://doi.org/10.3389/fmicb.2017.02224>
- GOLDBERG, C. S., TURNER, C. R., DEINER, K., KLYMUS, K. E., THOMSEN, P. F., MURPHY, M. A., SPEAR, S. F., MCKEE, A., OYLER-MCCANCE, S. J., CORNMAN, R. S., LARAMIE, M. B., MAHON, A. R., LANCE, R. F., PILLIOD, D. S., STRICKLER, K. M., WAITS, L. P., FRIEMER, A. K., TAKAHARA, T., HERDER, J. E., TABERLET, P. (2016) Critical considerations for the application of environmental DNA methods to detect aquatic species. *Methods in Ecology and Evolution*, 7(11), 1299–1307. <https://doi.org/10.1111/2041-210X.12595>
- GRÜNING, B., DALE, R., SJÖDIN, A., CHAPMAN, B. A., ROWE, J., TOMKINS-TINCH, C. H., VALIERIS, R., KÖSTER, J. (2018) Bioconda: Sustainable and comprehensive software distribution for the life sciences. *Nature Methods*, 15(7), 475–476. <https://doi.org/10.1038/s41592-018-0046-7>
- HAMILTON, D. P., CAREY, C. C., ARVOLA, L., ARZBERGER, P., BREWER, C., COLE, J. J., GAISER, E., HANSON, P. C., IBELINGS, B. W., JENNINGS, E., KRATZ, T. K., LIN, F.-P., MCBRIDE, C. G., DE EYTO, E., WEATHERS, K. C. (2015) A Global Lake Ecological Observatory Network (GLEON) for synthesising high-frequency sensor data for validation of deterministic ecological models. *Inland Waters*, 5(1), 49–56. <https://doi.org/10.5268/IW-5.1.566>

- HAN, Y., HE, J., LI, M., PENG, Y., JIANG, H., ZHAO, J., LI, Y., DENG, F. (2024) Unlocking the potential of metagenomics with the PacBio high-fidelity sequencing technology. *Microorganisms*, 12(12), 2482. <https://doi.org/10.3390/microorganisms12122482>
- HATVANI, I. G., KOVÁCS, J., SZÉKELY KOVÁCS, I., JAKUSCH, P., KORPONAI, J. (2011) Analysis of long-term water quality changes in the Kis-Balaton Water Protection System with time series-, cluster analysis and Wilks' lambda distribution. *Ecological Engineering*, 37(4), 629–635. <https://doi.org/10.1016/j.ecoleng.2010.12.028>
- HATVANI, I. G., DE BARROS, V. D., TANOS, P., KOVÁCS, J., SZÉKELY KOVÁCS, I., CLEMENT, A. (2020) Spatiotemporal changes and drivers of trophic status over three decades in the largest shallow lake in Central Europe, Lake Balaton. *Ecological Engineering*, 151, 105861. <https://doi.org/10.1016/j.ecoleng.2020.105861>
- HUGENHOLTZ, P., GOEBEL, B. M., PACE, N. R. (1998) Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *Journal of Bacteriology*, 180(18), 4765–4774. <https://doi.org/10.1128/JB.180.18.4765-4774.1998>
- ISTVÁNOVICS, V., CLEMENT, A., SOMLYÓDY, L., SPECZIÁR, A., G-TÓTH, L., PADISÁK, J. (2007) Updating water quality targets for shallow Lake Balaton (Hungary), recovering from eutrophication. In B. Qin, Z. Liu, & K. Havens (Eds.), *Eutrophication of Shallow Lakes with Special Reference to Lake Taihu, China*. Springer. pp. 305–318. https://doi.org/10.1007/978-1-4020-6158-5_31
- J. CRAIG VENTER INSTITUTE (2004) Global Ocean Sampling Expedition (GOS). <https://www.jcvi.org/research/gos>
- JOHNSON, J. S., SPAKOWICZ, D. J., HONG, B.-Y., PETERSEN, L. M., DEMKOWICZ, P., CHEN, L., ... WEINSTOCK, G. M. (2019) Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nature Communications*, 10, 5029. <https://doi.org/10.1038/s41467-019-13036-1>
- JOHNSON, S. S., ZAIKOVA, E., GOERLITZ, D. S., BAI, Y., TIGHE, S. W. (2017) Real-time DNA sequencing in the Antarctic Dry Valleys using the Oxford Nanopore sequencer. *Journal of Biomolecular Techniques*, 28(1), 2–7. <https://doi.org/10.7171/jbt.17-2801-009>
- KIM, D., SONG, L., BREITWIESER, F. P., SALZBERG, S. L. (2016) Centrifuge: Rapid and sensitive classification of metagenomic sequences. *Genome Research*, 26(12), 1721–1729. <https://doi.org/10.1101/gr.210641.116>
- KIM, M., PONGPANICH, M., PORNTAVEETUS, T. (2024) Unraveling metagenomics through long-read sequencing: A comprehensive review. *Journal of Translational Medicine*, 22, 111. <https://doi.org/10.1186/s12967-024-04917-1>
- KNIGHT, R., JANSSON, J., FIELD, D., FIERER, N., DESAI, N., FUHRMAN, J. A., ... GILBERT, J. A. (2012) Unlocking the potential of metagenomics through replicated experimental design. *Nature Biotechnology*, 30(6), 513–520. <https://doi.org/10.1038/nbt.2235>
- KOPF, A., BICAK, M., KOTTMANN, R., SCHNETZER, J., KOSTADINOV, I., LEHMANN, K., ... GLÖCKNER, F. O. (2015) The Ocean Sampling Day Consortium. *Standards in Genomic Sciences*, 10, 27. <https://doi.org/10.1186/s13742-015-0066-5>
- KOVÁCS, J., NAGY, M., CZAUNER, B., SZÉKELY KOVÁCS, I., BORSODI, A. K., HATVANI, I. G. (2012) Delimiting sub-areas in water bodies using multivariate data analysis on the example of Lake Balaton (W Hungary). *Journal of Environmental Management*, 110, 151–158. <https://doi.org/10.1016/j.jenvman.2012.06.002>
- KÖSTER, J., RAHMANN, S. (2012) Snakemake—A scalable bioinformatics workflow engine. *Bioinformatics*, 28(19), 2520–2522. <https://doi.org/10.1093/bioinformatics/bts480>
- LANGILLE, M. G. I., ZANEVELD, J., CAPORASO, J. G., MCDONALD, D., KNIGHTS, D., REYES, J. A., ... BEIKO, R. G. (2013) Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature Biotechnology*, 31(9), 814–821. <https://doi.org/10.1038/nbt.2676>
- LI, H. (2018) Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18), 3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>
- LIN, X., WARING, K., GHEZZI, H., TROPINI, C., TYSON, J., ZIELS, R. M. (2024) High accuracy meets high throughput for near full-length 16S ribosomal RNA amplicon sequencing on the Nanopore platform. *PNAS Nexus*, 3(10), pgae411. <https://doi.org/10.1093/pnasnexus/pgae411>
- LINZ, A. M., AYLWARD, F. O., BERTILSSON, S., MCMAHON, K. D. (2020) Time-series metatranscriptomes reveal conserved patterns between phototrophic and heterotrophic microbes in diverse freshwater systems. *Limnology and Oceanography*, 65(8), 1767–1783. <https://doi.org/10.1002/lno.11306>
- LOZUPONE, C., KNIGHT, R. (2005) UniFrac: A new phylogenetic method for comparing microbial communities. *Applied and Environmental Microbiology*, 71(12), 8228–8235. <https://doi.org/10.1128/AEM.71.12.8228-8235.2005>
- MANDAL, S., VAN TREUREN, W., WHITE, R. A., EGGESBØ, M., KNIGHT, R., PEDDADA, S. D. (2015) Analysis of composition of microbiomes: A novel method for studying microbial composition. *Microbial Ecology in Health and Disease*, 26, 27663. <https://doi.org/10.3402/mehd.v26.27663>
- MARRONE, B. L., BANERJEE, S., TALAPATRA, A., GONZALEZ-ESQUER, C. R., PILANIA, G. (2023) Toward a predictive understanding of cyanobacterial harmful algal blooms through AI integration of physical, chemical, and biological data. *ACS ES&T Water*, 4, 844–858. <https://doi.org/10.1021/acsestwater.3c00369>
- McMURDIE, P. J., HOLMES, S. (2014) Waste not, want not: Why rarefying microbiome data is inadmissible. *PLOS Computational Biology*, 10(4), e1003531.

- <https://doi.org/10.1371/journal.pcbi.1003531>
- MENEGON, M., CANTALONI, C., RODRIGUEZ-
PRIETO, A., CENTOMO, C., ABDELFATTAH, A.,
ROSSATO, M., ... DELLEDONNE, M. (2017) On site
DNA barcoding by nanopore sequencing. *PLOS ONE*,
12(10), e0184741. <https://doi.org/10.1371/journal.pone.0184741>
- NATIONAL ECOLOGICAL OBSERVATORY
NETWORK (NEON) (2024) Microbial community
sampling protocol (NEON.DOC.015675). Available at:
<https://www.neonscience.org/>
- NEARING, J. T., DOUGLAS, G. M., HAYES, M. G.,
MACDONALD, J., DESAI, D. K., ALLWARD, N., ...
LANGILLE, M. G. I. (2022) Microbiome differential
abundance methods produce different results across 38
datasets. *Nature Communications*, 13, 342. <https://doi.org/10.1038/s41467-022-28034-z>
- NEWTON, R. J., JONES, S. E., EILER, A., MCMAHON,
K. D., BERTILSSON, S. (2011) A guide to the natural
history of freshwater lake bacteria. *Microbiology and
Molecular Biology Reviews*, 75(1), 14–49. <https://doi.org/10.1128/MMBR.00028-10>
- OLEKSYK, T. Kh., WOLFSBERGER, V. V.,
SHCHUBELKA, Kh. M., HASYNETS, Y. S., KISH,
R. Y., VAKERYCH, M. M., MIRUTENKO,
V. S., HUKLYVSKA, A. V., KURUTSA, O. S.,
DUBIV, O. V., SHYLOVA, N. O., KEDYK, H. V.,
SHTEFANIAK, I. I. (2025) Khab na kordoni: UzhNU
yak tsentr bioinformatyky ta henomiky v umovakh
viiny. *Proceedings of the 79th Conference of
professors and lecturers of UzhNU, Seriya Biolohiia*,
Tom I. 27.02.2025, Uzhhorod, Vydavnytstvo UzhNU
«Hoverla», p. 19. (in Ukrainian)
- PACE, N. R. (1997) A molecular view of microbial diversity
and the biosphere. *Science*, 276(5313), 734–740.
<https://doi.org/10.1126/science.276.5313.734>
- QUAST, C., PRUESSE, E., YILMAZ, P., GERKEN,
J., SCHWEER, T., YARZA, P., ... GLÖCKNER,
F. O. (2013) The SILVA ribosomal RNA gene database
project: Improved data processing and web-based tools.
Nucleic Acids Research, 41(D1), D590–D596. <https://doi.org/10.1093/nar/gks1219>
- QUICK, J., LOMAN, N. J., DURAFFOUR, S., SIMPSON,
J. T., SEVERI, E., COWLEY, L., ... CARROLL,
M. W. (2016) Real-time, portable genome sequencing
for Ebola surveillance. *Nature*, 530(7589), 228–232.
<https://doi.org/10.1038/nature16996>
- QUINCE, C., WALKER, A. W., SIMPSON, J. T., LOMAN,
N. J., SEGATA, N. (2017) Shotgun metagenomics,
from sampling to analysis. *Nature Biotechnology*,
35(9), 833–844. <https://doi.org/10.1038/nbt.3935>
- QUINN, T. P., ERB, I., GLOOR, G. B., NOTREDAME,
C., RICHARDSON, M. F., CROWLEY, T. M. (2019)
A field guide for the compositional analysis of any-
omics data. *GigaScience*, 8(9), giz107. <https://doi.org/10.1093/gigascience/giz107>
- RUSCH, D. B., HALPERN, A. L., SUTTON, G.,
HEIDELBERG, K. B., WILLIAMSON, S., YOOSEPH,
S., ... VENTER, J. C. (2007) The Sorcerer II Global
Ocean Sampling expedition: Northwest Atlantic
through Eastern Tropical Pacific. *PLOS Biology*, 5(3),
e77. <https://doi.org/10.1371/journal.pbio.0050077>
- SCHRADER, C., SCHIELKE, A., ELLERBROEK, L.,
& JOHNE, R. (2012) PCR inhibitors—Occurrence,
properties and removal. *Journal of Applied
Microbiology*, 113(5), 1014–1026. <https://doi.org/10.1111/j.1365-2672.2012.05384.x>
- SUNAGAWA, S., COELHO, L. P., CHAFFRON, S.,
KULTIMA, J. R., LABADIE, K., SALAZAR, G.,
... BORK, P. (2015) Ocean plankton: Structure and
function of the global ocean microbiome. *Science*,
348(6237), 1261359. <https://doi.org/10.1126/science.1261359>
- TÁTRAI, I., MÁTYÁS, K., KORPONAI, J., PAULOVITS,
G., POMOGYI, P. (2000) The role of the Kis-Balaton
Water Protection System in the control of water quality
of Lake Balaton. *Ecological Engineering*, 16(1), 73–78.
[https://doi.org/10.1016/S0925-8574\(00\)00091-4](https://doi.org/10.1016/S0925-8574(00)00091-4)
- THOMPSON, L. R., SANDERS, J. G., MCDONALD, D.,
AMIR, A., LADAU, J., LOCEY, K. J., PRILL, R. J.,
TRIPATHI, A., GIBBONS, S. M., ACKERMANN, G.,
NAVAS-MOLINA, J. A., JANSSEN, S., KOPYLOVA,
E., VÁZQUEZ-BAEZA, Y., GONZÁLEZ, A.,
MORTON, J. T., MIRARAB, S., ZECH XU, Z.,
JIANG, L., ... KNIGHT, R. (2017) A communal
catalogue reveals Earth’s multiscale microbial diversity.
Nature, 551(7681), 457–463. <https://doi.org/10.1038/nature24621>
- TÓTH, V. R. (2023) Exploring the functional connectivity
between the Kis-Balaton Water Protection System and
Lake Balaton using satellite data. *Aquatic Sciences*, 85,
4. <https://doi.org/10.1007/s00027-023-01005-2>
- TÓTH, V. R., CASTRO-MARQUEZ, S., WOLFSBERGER,
W., OLEKSYK, T. K. (2025) Effect of habitat type on
the bacterioplankton communities in a semi-artificial
wetland. Manuscript submitted for publication.
- URBAN, L., HOLZER, A., BARONAS, J. J., HALL,
M. B., BRAEUNINGER-WEIMER, P., SCHERM,
M. J., ... PALLEN, M. J. (2021) Freshwater monitoring
by nanopore sequencing. *eLife*, 10, e61504. <https://doi.org/10.7554/eLife.61504>
- VANDEPUTTE, D., KATHAGEN, G., D’HOE, K.,
VIEIRA-SILVA, S., VALLES-COLOMER, M.,
SABINO, J., ... RAES, J. (2017) Quantitative
microbiome profiling links gut community variation
to microbial load. *Nature*, 551(7681), 507–511. <https://doi.org/10.1038/nature24460>
- VENTER, J. C., REMINGTON, K., HEIDELBERG,
J. F., HALPERN, A. L., RUSCH, D., EISEN, J. A., ...
SMITH, H. O. (2004) Environmental genome shotgun
sequencing of the Sargasso Sea. *Science*, 304(5667),
66–74. <https://doi.org/10.1126/science.1093857>
- WEISS, S., XU, Z. Z., PEDDADA, S., AMIR, A.,
BITTINGER, K., GONZALEZ, A., ... KNIGHT,
R. (2017) Normalization and microbial differential
abundance strategies depend upon data characteristics.
Microbiome, 5, 27. <https://doi.org/10.1186/s40168-017-0237-y>

- WIDDER, S., ALLEN, R. J., PFEIFFER, T., CURTIS, T. P., WIUF, C., SLOAN, W. T., ... KLAPPER, I. (2016) Challenges in microbial ecology: Building predictive understanding of community function and dynamics. *The ISME Journal*, 10(11), 2557–2568. <https://doi.org/10.1038/ismej.2016.45>
- WILKINSON, M. D., DUMONTIER, M., AALBERSBERG, I. J., APPLETON, G., AXTON, M., BAAK, A., ... MONS, B. (2016) The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>
- WOESE, C. R., FOX, G. E. (1977) Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proceedings of the National Academy of Sciences*, 74(11), 5088–5090. <https://doi.org/10.1073/pnas.74.11.5088>
- WOOD, D. E., LU, J., LANGMEAD, B. (2019) Improved metagenomic analysis with Kraken 2. *Genome Biology*, 20, 257. <https://doi.org/10.1186/s13059-019-1891-0>
- YILMAZ, P., KOTTMANN, R., FIELD, D., KNIGHT, R., COLE, J. R., AMARAL-ZETTLER, L., GILBERT, J. A., KARSCH-MIZRACHI, I., JOHNSTON, A., COCHRANE, G., VAUGHAN, R., HUNTER, C., PARK, J., MORRISON, N., ROCCA-SERRA, P., STERK, P., ARUMUGAM, M., BAILEY, M., BAUMGARTNER, L., ... GLÖCKNER, F. O. (2011) Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIXS) specifications. *Nature Biotechnology*, 29(5), 415–420. <https://doi.org/10.1038/nbt.1823>
- YOOSEPH, S., SUTTON, G., RUSCH, D. B., HALPERN, A. L., WILLIAMSON, S. J., REMINGTON, K., ... VENTER, J. C. (2007) The Sorcerer II Global Ocean Sampling expedition: Expanding the universe of protein families. *PLOS Biology*, 5(3), e16. <https://doi.org/10.1371/journal.pbio.0050016>
- ZHANG, T., LI, H., MA, S., CAO, J., LIAO, H., HUANG, Q., CHEN, W. (2023) The newest Oxford Nanopore R10.4.1 full-length 16S rRNA sequencing enables the accurate resolution of species-level microbial community profiling. *Applied and Environmental Microbiology*, 89(10), e00605-23. <https://doi.org/10.1128/aem.00605-23>

Дата першого надходження статті до видання: 01.03.2026

Дата прийняття статті до друку після рецензування: 08.04.2026

Дата публікації (оприлюднення) статті: 29.05.2026